

# Grounding physicalism and the knowledge argument

---

Alex Moran,  
Université de Fribourg<sup>1</sup>

This paper is forthcoming in *Philosophical Perspectives*.

**ABSTRACT:** Standard responses to the knowledge argument grant that Mary could know all of the physical facts even while trapped inside her black-and-white room. What they deny is that upon leaving her black-and-white room and experiencing red for the first time, Mary learns a genuinely new fact. This paper develops an alternative response in a grounding physicalist framework, on which Mary does not know all of the physical facts while trapped inside the room. The main thesis advocated is that Mary does not know certain phenomenal facts while trapped inside the room, whereby these facts classify as physical due to being wholly and fully metaphysically grounded in the underlying fundamental facts which are themselves entirely physical.

[The knowledge argument] implies only that some facts about colour experiences cannot be learned by watching lectures on black-and-white television... This, however, does not yet license any further conclusions about the nature of [these] experiences. In particular, it does not entitle us to infer that these experiences are not physical events.

— Alter, T. (1998: 50)

## 1 Introduction

One of the main arguments against physicalism is the ‘knowledge argument’, due originally to Frank Jackson (1982).<sup>2</sup> The argument turns on a familiar thought experiment involving Mary, a ‘brilliant scientist forced to investigate the world from a black and white room’ (ibid: 30). Whilst inside the room, the story goes, Mary learns all of the physical information involved in experiencing red. Then, one day, she leaves the room and experiences red for the first time. Plausibly, Mary thereby learns a *new* fact, concerning what red experiences are qualitatively like. However, since Mary knew all the relevant physical facts before leaving the room, it follows that the new fact that she learns must be non-physical. But this means that physicalism – construed, plausibly, as entailing that all the facts are physical – is false.

In short, the knowledge argument, if it is sound, establishes that physicalism leaves something out. As Jackson (1986: 291) explains: “[P]hysicalists must hold that complete physical knowledge is complete knowledge simpliciter. It seems,

---

<sup>1</sup> Versions of this paper have been presented to audiences in Warsaw, Grenoble, Bochum, London, Fribourg, Stockholm, Oxford, Barcelona, Southampton, and Trinity College Dublin. I’d like to thank the participants on these occasions for their valuable comments. Thanks also to Ralf Bader, David Chalmers, Dominic Alford-Duguid, John Divers, Anandi Hattiangadi, Mark Johnston, James Openshaw, Jonathan Schaffer, Mikołaj Slawkowski-Rode, and Ralph Weir. This research was supported by a grant from the Irish Research Council, whose support I gratefully acknowledge.

<sup>2</sup> A version of the argument was also articulated at the same time by Robinson (1982). Early expressions of what is arguably the same idea appear in Broad (1925) and Russell (1927).

however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red...Hence, physicalism is false.”

Given the simplifying assumption (which is harmless in the present context) that if Mary knows all the relevant physical facts about experiencing red (and so on) whilst still inside the room, then at that point she knows all of the physical facts simpliciter, the knowledge argument can be formalised as follows:

- |       |  |                |
|-------|--|----------------|
| 1     | In her black and white room, Mary knows all the physical facts.      |                |
| 2     | When Mary experiences red for the first time, she learns a new fact. |                |
| <hr/> |  |                |
| ∴     | The new fact Mary learns is non-physical.                            | [from 1, 2]    |
| 3     | If physicalism is true, then all the facts are physical facts.       |                |
| <hr/> |  |                |
| ∴     | Physicalism is false.  | [from 1, 2, 3] |

A great deal has been written about the knowledge argument over the years. Indeed, it is probably one of the most discussed arguments in the vast contemporary literature on the mind-body problem. Notably, however, whilst physicalists have defended a broad range of responses to the argument, there is a general consensus regarding the basic strategy that physicalists must adopt. In particular, it is widely assumed that to adequately respond to the knowledge argument, physicalists must accept the first premise, and focus their efforts on denying the second. In other words, the consensus is that physicalists should grant that Mary knows all the physical facts whilst still inside the room, but then deny, in one way or another, that Mary learns a genuinely new fact upon experiencing red for the first time.

However, precisely by virtue of being instances of this common strategy, extant physicalist responses to the knowledge argument also share a common defect, namely insofar as each is incompatible with the powerful intuition that Mary learns something genuinely new when she experiences red for the first time. As for the intuition itself, I have no new arguments in its favour. But, like many, I find it overwhelmingly plausible to think that when Mary first experiences red, she thereby learns a new fact concerning what red experiences are like.<sup>3</sup> Accordingly, I am inclined to agree with Jackson’s assessment in the following passage:

What will happen when Mary is released from her black-and-white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. (1982: 130)

Where I part ways with Jackson, though, is when it comes to the idea that accepting this intuition forces us to abandon physicalism. Again, this is not only Jackson’s position, but also the consensus view in the surrounding literature. My own view, however, is that one can accept this intuition even within a robustly physicalist framework. The goal of this paper is to elaborate and defend this claim.<sup>4</sup>

---

<sup>3</sup> It is perhaps worth noting here a result from recent work in experimental philosophy. When asked about the Mary case, 80% of respondents to a survey agreed that Mary did not know what it was like to see red until she left the room and had a red experience. See Gregory *et al.* (2022).

<sup>4</sup> It is true that Jackson later abandoned the view that the knowledge argument undermines physicalism (see Jackson 2003). However, it is unclear that Jackson ever abandoned the conditional claim that if indeed Mary learns something genuinely new, then physicalism is false. (What Jackson

My strategy will be to operate in a framework that captures the doctrine of physicalism in terms of the notion of ‘metaphysical grounding’. On this view, the core physicalist claim is that mental facts are metaphysically grounded in underlying physical facts; or, in other words, that mental facts obtain in virtue of underlying physical facts. What I shall argue is that operating in this framework enables one to plausibly maintain that whilst Mary learns a genuinely new fact concerning what red experience is like, this *phenomenal* fact is also a *physical* fact. We can then claim that the trouble with the knowledge argument is not the second premise, i.e., that Mary learns a genuinely new fact, but rather the first premise, i.e., that Mary knows all of the physical facts even before leaving the black-and-white room. In this way, we will be able to preserve the powerful intuition that Mary learns a genuinely new fact upon leaving the room even within a physicalist framework, thus undermining then commonly held view that this intuition is the sole preserve of non-physicalist theories (cf. Goff 2017; Nida-Rümelin 2007; Robinson 2016).

The paper is structured as follows. §2 elaborates on the sense in which I take Mary to learn something genuinely new, and also clarifies the way in which the present view differs from extant responses in the literature. §3 sets out the grounding physicalist response to the knowledge argument that I want to defend. §§4-5 develop this response further in relation to some objections. §6 concludes.

## 2 What does Mary learn?

What does Mary learn when she comes out of the room and experiences red for the first time? Plausibly, she learns a new, *phenomenal* fact concerning the qualitative character of red experiences. That is, she learns *what it is like* to experience red. (Roughly, this is the fact that Mary might express with the words ‘This is what red experiences are like’.) Let us refer, in what follows, to this phenomenal fact as ‘ $\Delta$ ’.

For present purposes, I want to think of facts as being worldly entities, roughly as Wittgenstein (1922) does in the *Tractatus*. The idea is that to know a fact (in this sense) is to stand in a distinctive cognitive relation to a portion of reality, whereby such portions of reality might be thought of as property instantiations or states of affairs (in the sense of Armstrong 1997).<sup>5</sup> Of course, one can also think of facts as representational entities, namely as true propositions, and one can just as well construe these as being the objects of knowledge (i.e. the things known). Nor do I want to deny that knowing a fact in the first sense also involves entertaining a true proposition (or having a true Fregean thought) and hence knowing a fact in the second sense.<sup>6</sup> Rather, I want to operate with something like the following picture: knowing a worldly fact involves entertaining a true proposition that represents that fact and that is isomorphic to it in some appropriate way.<sup>7</sup> One can

---

later came to deny, by means of drawing on a representationalist approach to the nature of consciousness, is that Mary does indeed learn anything new in the sense that is at issues in this paper.)

<sup>5</sup> The conception of facts developed in Fine (1982) would also suit our purposes.

<sup>6</sup> Frege (1918-19) of course maintained that facts *just are* thoughts that are true. The Fregean conception of facts as representational media thus contrasts with the Wittgensteinian conception of facts as portions of reality (cf. Crane 2019: 21).

<sup>7</sup> For instance, the fact might be a concrete state of affairs consisting in an object instantiating a property, whereas the true proposition that represents that fact might be a structured item that

then think of the relation between the proposition entertained (the Fregean thought) and the fact known (i.e. the worldly fact or state of affairs) in terms of reference, whereby the role of the proposition (or thought) is to represent the world, such that when the proposition is true it has a worldly fact as its referent.

In this framework, we can think of  $\Delta$  as being a worldly item consisting in the instantiation of some phenomenal property  $P$  by a certain experience-type  $\Phi$  (namely experiences of the sort we naturally refer to as ‘red experiences’). One main claim that I want to advocate can thus be understood as the idea that Mary learns something new in the precise sense that she learns a new phenomenal fact (conceived as an aspect of the world) that she did not already know. On this view, there was a part of reality that Mary did not stand in the cognitive relation of ‘knowing’ to until she left the room and experienced of red. Accordingly, the view can be distinguished from a range of alternate responses on which (in some sense at least) Mary learns something new when she escapes her achromatic chamber.

On one alternate approach, for example, while Mary learns something new after first experiencing red, she does not thereby learn a new fact. Rather, she gains some other sort of knowledge – perhaps *knowledge-how* (abilities-knowledge) or *knowledge by acquaintance*.<sup>8</sup> On yet another approach, while Mary does learn a new fact, this merely involves coming to entertain a new proposition that represents an already familiar portion of reality in a way that was previously unavailable to her.<sup>9</sup> Both approaches differ in important ways from the view advocated here. On the view I will defend, Mary does not merely gain new know-how or acquaintance-knowledge; rather she learns a genuinely new fact about the world. And in so doing, Mary does not merely come to represent information that she already knows in a new way; rather, she comes to know a brand new worldly fact, and, hence, becomes acquainted with a portion of reality that she was not already familiar with.

The contrast between this second (old information/new fact) approach and my own view can be further highlighted as follows. On the old information/new fact view there is some worldly fact  $\Omega$  that Mary knows inside the room by means of entertaining some Fregean thought  $T_1$ . The idea then is that upon leaving the room, Mary comes to entertain some new Fregean thought  $T_2$  which also has  $\Omega$  as its referent. Accordingly, Mary does learn a new fact – in the sense of entertaining a new (true) Fregean thought  $T_2$ . But this does not involve gaining knowledge of a new worldly fact, for Mary already knew all the worldly facts inside the room. On my approach, by contrast, Mary does not merely come to entertain a new (true) Fregean thought after experiencing red. Rather, she comes to have knowledge of a new worldly fact,  $\Delta$ . The difference between Mary before she leaves the room and Mary afterwards is therefore not merely a difference at the level of representation (concerning the ways in which she is able to represent the world), but rather a difference at the level of facts (concerning the cognitive relations she bears to parts

---

binds the object to the property by means of the relation of predication (e.g. a true Russellian proposition). For a detailed and compelling development of this sort of picture see Johnston (2006).

<sup>8</sup> For the first (ability-theoretic) approach see Lewis (1988) and Nemirow (1980); for the latter (acquaintance-theoretic) approach see Connee (1994) and Tye (2009).

<sup>9</sup> This kind of view is typically developed as part of a broader ‘phenomenal concepts strategy’ for handling the problems of consciousness (cf. Stoljar 2005, Nida-Rümelin & O’Conaill 2021). However, views of this kind can also be developed without bringing in phenomenal concepts (cf. Crane 2003, 2019; Perry 2001). I return to the issue of phenomenal concepts briefly below in §4.

of the world itself). Before leaving the room, there was a portion of the world that Mary did not stand in the relation of ‘knowing’ to (namely the phenomenal fact  $\Delta$ ), and she came to stand in that relation to that portion of reality only after first experiencing red.

The question we must answer now concerns the sense in which  $\Delta$  could nonetheless be a physical fact. In the next section, I develop an answer to this question. The main idea is that  $\Delta$  classifies as a genuinely physical fact by virtue of being metaphysically grounded in underlying fundamental physical facts.

### 3 A grounding physicalist response

The first premise of the knowledge argument states that even inside the black-and-white room, and prior to experiencing red, Mary knew all of the physical facts. Many in the literature accept this premise, including both physicalists and their opponents. However, premise 1 is not obviously true. In fact, I think it both can and should be rejected, on the grounds that there are some physical facts that Mary did not know whilst still inside the room, including, crucially, the phenomenal fact she learns upon leaving the room and experiencing red for the first time.

Why think that premise (1) is true? Jackson himself offers little by way of positive argument. Rather, he seems to think that the premise is obvious:

It can hardly be denied that it is in principle possible to obtain all [the relevant] physical information from black and white television, otherwise the Open University would of necessity need to use colour television (Jackson 1982: 30).

By itself, however, this passage is insufficient to establish that Mary could have known all the physical facts there are to know inside the room. The question, then, is what positive reasons can be given for accepting that premise.

One natural thought is that premise 1 might fall out of a plausible conception of what the physical facts actually are. For instance, one might think that the physical facts are just those facts that are or would be disclosed by a completed basic physics. Plausibly, on this conception of the physical, Mary could indeed learn all the physical facts inside the room, simply by means of black-and-white television and/or textbooks printed only in black-and-white. In fact, this is arguably one of the main insights contained in Bertrand Russell’s proto-knowledge argument (which appeared almost half a century before Jackson’s version). Russell writes:

It is obvious that a man who can see knows things which a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge which other men have and he has not is not a part of physics. (1927: 389)

Even if blind, Russell rightly maintains, one could in principle learn the entirety of physics (cf. Broad 1925: 71). In the same way, therefore, Mary could well learn the entirety of physics even from inside the black-and-white room.

Importantly, however, it remains open to physicalists to deny that the physical facts are just those facts disclosed by basic physics. Austere physicalists, who think that reality is exhausted by the realm of basic physics, may want to accept this picture. However, there is space for the more liberal sorts of physicalist to resist this

austere picture, in part at least by insisting that the class of physical facts is not exhausted by the facts disclosed by basic physics. In turn, this opens up space for such physicalists to resist the knowledge argument at its first step, by rejecting the assumption that Mary could know all the physical facts even in the room.<sup>10</sup>

To that end, consider a form of physicalism that is much more liberal than the austere kind of physicalism gestured at above, on which reality can be described at many different levels, and hence is not exhausted by the ontology of basic physics. On such a view, basic physics does not describe reality in total, but rather only the fundamental level that constitutes the ‘metaphysical ground floor’ (to take a helpful phrase from Bennett 2001: 76). Reality itself, meanwhile, comprises in addition a wide range of derivative items besides the sparse range of fundamental items recognised by basic physics. On one popular version of this picture, for example, the fundamental level comprises the mereologically simple items and their basic properties, and the various derivative levels are then populated by increasingly complex items of various kinds, exhibiting various (derivative) properties and relations.<sup>11</sup>

There is room for disagreement as to how to best flesh out this sort of liberal picture. My own view is that the best way to do so is by employing the neo-Aristotelian notion of metaphysical grounding, conceived as a primitive relation of metaphysical dependence and determination that is paradigmatically expressed by means of the ‘in virtue of’ locution (see, among others, Audi 2012; Fine 2012; Rosen 2010; Schaffer 2009). On this picture, the fundamental level of reality is the level of basic physics, and all other phenomena, including mental phenomena, are metaphysically grounded therein. Or to state the view using the language of facts, the idea is that all the fundamental facts are those of basic physics, and that all other facts, including mental facts, are derivative facts that obtain in virtue of the fundamental physical ones. The fact that I am now in pain, for example, obtains in virtue of, and is ultimately grounded in, some suitable array of basic physical facts.

This sort of physicalist view has recently been articulated by various philosophers either sympathetic to physicalism itself, or at least to the idea that the grounding apparatus can be fruitfully employed to capture an important version of that view (cf. Bader manuscript; Dasgupta 2014; Goff 2017; Lui forthcoming, Moran 2021, 2023; Schaffer 2017; Trogon 2017). In this paper, I will assume that grounding physicalism is a genuine and interesting kind of physicalism, without arguing separately for that claim. The aim will rather be to demonstrate the work that can be done on that assumption, by means of developing a novel response to the knowledge argument.

Consider again the phenomenal fact  $\Delta$  that Mary learns upon first experiencing red. In a grounding physicalist framework,  $\Delta$  will be a derivative physical fact that is metaphysically grounded in some plurality of fundamental physical facts. Call this plurality of fundamental physical grounding facts  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ . (These facts comprise the ultimate metaphysical grounds of  $\Delta$ .) Grounding physicalism implies:

---

<sup>10</sup> Cf. Montero (2007: 178), who emphasises a very similar point, in the course of developing an answer to the knowledge argument that in many ways anticipates the one developed here.

<sup>11</sup> For helpful discussion of this sort of levels-based ontology see Kim (1998, 2002) and Oppenheim & Putnam (1956). Cf. also deRosset (forthcoming) and Schaffer (2010). For helpful discussion of the contrast between austere and liberal forms of physicalism see Pautz (forthcoming).

(G)  $\Delta < \Gamma_1, \Gamma_2, \dots, \Gamma_n$

That is:  $\Delta$  is metaphysically grounded in – obtains in virtue of –  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ .

My central claim is that consistently with all of this, we can maintain that Mary did not know  $\Delta$  while still inside the room, and that she learned  $\Delta$  only after escaping the room and undergoing a red experience. On a grounding physicalist view, we can recognise that there exist facts that do not reduce to lower-level physical facts but that are nonetheless grounded therein.<sup>12</sup> We can then say that such facts, while not being *narrowly* physical facts, i.e. while not being among the basic physical facts at the fundamental level, are nonetheless *broadly* physical facts, by virtue of being wholly and fully grounded<sup>13</sup> in the narrowly physical facts comprising the metaphysical ground floor.<sup>14</sup> The grounding physicalist can also insist that phenomenal facts like  $\Delta$  are such broadly physical facts, and so reject the knowledge argument by denying premise 1, on the grounds that while Mary might have known all of the narrowly physical facts, it does not follow that she knew all of the physical facts *simpliciter*, since there was at least one broadly physical fact she did not know, viz. the phenomenal fact she learns upon leaving the room.<sup>15</sup>

Of course, there are other liberal (non-austere) forms of physicalism besides grounding physicalism. For many philosophers, I suspect that supervenience physicalism will spring most naturally to mind. However, it is worth noting that is not at all obvious that supervenience is the right relation to appeal to in the present context. The reason why is that we need to be able to offer a robust sense in which the phenomenal fact  $\Delta$  that Mary learns is a genuinely physical fact. Yet, it is unclear that merely stating that  $\Delta$  supervenes on the basic physical facts is sufficient to achieve this. The core idea of supervenience is that there are certain physical facts that fix – that is to say, that metaphysically necessitate – the mental facts. A supervenience physicalist might therefore offer the following necessitation thesis:

---

<sup>12</sup> I don't mean here to take too firm a stance on the interesting and important question as to how exactly grounding relates to reduction. (Perhaps in some sense grounded facts are reducible to the facts that ground them; perhaps not.) For relevant discussion see Bader (manuscript), Dorsey (2016); Goff (2017); Rosen (2010).

<sup>13</sup> *Fully* in the sense of not merely partially grounding in underlying fundamental physical facts; *wholly* in the sense of not being grounded by any other (non-physical) facts.

<sup>14</sup> Cf. Schaffer (manuscript), who draws a similar distinction between physical and material facts (which corresponds to my distinction above between narrow and broad physical facts).

<sup>15</sup> Note briefly how the present view contrasts with the 'inclusive' form of 'subjective physicalism' described by Howell (2007, 2008, 2009). On that view, all of the objects and properties in the world are referred to by (a completed) basic physics; it is just that some of these properties, namely those involved in conscious experience, cannot be completely or adequately described without using concepts that can be made available to us solely by undergoing experience (see esp. Howell 2009). On the present view, by contrast, the objects and properties described by basic physics do not mention all of the objects and properties that there are. Rather, phenomenal properties are conceived as genuinely higher-level properties that do not show up in a completed fundamental physics and that are instantiated by items (namely conscious experiences) that also do not show up in any such physics but that are located instead only at higher, derivative levels of being. (Another important contrast is that I am not inclined to grant that there are physical facts knowable *only* on the basis of conscious experience; hence, the view on offer here also differs from the 'exclusive subjective physicalism' that Howell discusses. Cf. §4 below for further discussion.)

(S)  $\Box (\Gamma_1, \Gamma_2, \dots, \Gamma_n \supset \Delta)$

Essentially, this tells us that once you fix the relevant fundamental physical facts  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ , you're bound to get the phenomenal fact  $\Delta$ .<sup>16</sup> So, "all God had to do", to employ Kripke's familiar metaphor, in order to ensure that the phenomenal fact should hold, is to ensure that the relevant physical facts should hold. What I want to draw into question is whether this thesis should really be taken to provide sufficient materials for believing that  $\Delta$  is a genuinely physical fact. Consider a relevantly similar view in meta-ethics, according to which normative properties are wholly distinct (not just numerically but also in kind) from natural properties. It is common for non-naturalists who accept this kind of view to maintain that, nevertheless, the non-natural normative properties strongly supervene upon the natural non-normative properties. Indeed, this is precisely the kind of non-naturalist view that G. E. Moore advocates in the *Principia Ethica* (Moore 1903). What consideration of this theory shows, I submit, is that in principle at least, supervenience relations could hold between entirely distinct families of properties. The normative properties might be *sui generis*, and not at all dependent on the descriptive properties, and yet a supervenience relation might nevertheless obtain.<sup>17</sup> Moreover, it seems that we could easily imagine an analogous coherent dualist position on which, while mental properties and physical properties are wholly distinct features of different kinds, nevertheless, mental properties supervene upon physical properties. Such a dualist might accept (S), but would also insist that  $\Delta$  is non-physical. I want to say that these reflections undermine the thought that one can adequately capture the idea that  $\Delta$  is physical by appealing only to (S).<sup>18,19</sup>

The grounding physicalist, by contrast, is not in this position. This is because the claim that mental facts like  $\Delta$  are grounded in fundamental physical facts

---

<sup>16</sup> Some might wish to interpret the box operator so that it ranges not over all metaphysically possible worlds but only certain such worlds, e.g. those that David Lewis referred to as the 'inner sphere' (Lewis 1994). For present purposes, we can prescind from complications of this kind.

<sup>17</sup> One might of course raise questions as to how precisely normative properties could supervene on non-normative properties if indeed these families of properties are wholly distinct (as on non-naturalism). One way to do this would be to insist that there can be no unexplained supervenience relations, but that no plausible explanation for the supervenience of the normative on the non-normative could be given in a non-naturalist framework. Cf. Hattiangadi (2018). All that matters for present purposes, however, is that the above non-naturalist view be metaphysically coherent. To achieve this, we might imagine that God acts as the ground of the supervenience relation between the *sui generis* normative properties and the non-normative properties on which they supervene. Perhaps we've good reasons for rejecting such a view, but reflecting on this kind of position, I want to say, is sufficient for establishing that just because one type of property supervenes upon another, it need not follow that the supervenient properties aren't *sui generis* and independent of the properties on which they supervene. (For a longer argument to this effect, see Horgan 1993.)

<sup>18</sup> A growing number of philosophers appear to recognise that the apparatus of supervenience is insufficient for capturing the doctrine of physicalism or the idea that mental facts are (at least broadly) physical (see e.g. Horgan 1993; Schaffer 2017). What is less often appreciated is that a thesis of mental/physical supervenience might well be held in common between dualists and physicalists. This same conclusion is reached in Bader (forthcoming), though via a different route.

<sup>19</sup> None of this, of course, undermines the plausible idea that some sort of supervenience claim is a necessary condition on physicalism being true. I'm claiming only that the supervenience of mental facts on physical ones is by itself sufficient to ensure then physicality of the mental.



plausibly ensures that mental facts like  $\Delta$  themselves are (at least broadly) physical.<sup>20</sup> Several things can be said in favour of this claim; I will confine myself to three.

First, it is widely held (and not implausibly) that grounded facts in some sense *consist in* the facts that ground them (cf. Fine 2012; Goff 2017: Ch. 1; Rosen 2010). Accordingly, grounding physicalism implies that phenomenal facts in some sense consist in the fundamental physical facts that serve as their ultimate grounds. Plausibly, however, if any fact  $X$  consists in some array of fundamental physical facts  $\Gamma$ , then  $X$  itself is (at least broadly) physical. Accordingly, in a grounding physicalist framework, we can insist that phenomenal facts like  $\Delta$  must be (at least broadly) physical, by virtue of consisting in their fundamental physical grounds.<sup>21</sup>

Second, it is widely maintained (again plausibly) that grounding relations serve to back a distinctive kind of metaphysical explanation.<sup>22</sup> Accordingly, the claim that phenomenal facts are grounded in fundamental physical facts ensures that a metaphysical explanation of the phenomenal facts can be given in terms of the basic physical facts comprising the ground floor of being. Plausibly, however, any facts that can be metaphysically explained in terms of the fundamental physical facts should be reckoned as being themselves (at least broadly) physical. So facts like  $\Delta$  will come out as (at least broadly) physical in a grounding physicalist framework, due to being metaphysically explainable in an appropriately physicalistic way.

Third, and finally, grounding physicalism serves to place phenomenal facts like  $\Delta$  right alongside various other manifestly (broadly) physical facts, such as the biological facts or chemical facts. In a grounding physicalist framework, all of these facts are derivative facts, ultimately grounded in the fundamental physical array. Since it is entirely reasonable to insist that the biological and the chemical facts (for instance) are (broadly) physical precisely by virtue of being grounded in the basic physical array, it also seems entirely reasonable to claim that phenomenal facts like  $\Delta$ , which are also grounded in that array, are (broadly) physical for the same reason.

Could one advance a similar line of thought in a different (non-ground-theoretic) liberal physicalist framework, say one that employs a relation like realisation, or constitution, rather than metaphysical grounding? Perhaps, and in my book, that would be all to the good. All that I want to insist here, however, is that one can at least develop the above line of response in a grounding physicalist setting. In that setting, we can insist that while Mary learns something new upon leaving the room, the fact she learns is a (broadly) physical fact, grounded (like all other broadly physical facts) in the fundamental physical facts that can be read off from basic physics. It then follows that the knowledge argument is unsound

---

<sup>20</sup> This is *pace* Nida-Rümelin & O’Conaill (forthcoming), who argue that dualism can be developed even in a grounding framework. I lack the space to engage with their arguments here. For further relevant discussion of this and related issues see Goff (2017: Ch. 1).

<sup>21</sup>Note, also, that if one did think that a supervenience thesis like (S) were sufficient for securing the (at least broad) physicality of  $\Delta$ , one should for this reason grant that the grounding physicalist thesis (G) is also sufficient for securing that result. This is because, given the plausible and widely held idea that grounds necessitate, ground physicalism implies (even if it is not implied by) supervenience physicalism, so that (G) entails (S) (although the converse does not hold.)

<sup>22</sup> Some philosophers hold that grounding is itself a relation of metaphysical explanation (so-called ‘unionists’), others hold rather that grounding is a relation apt to back metaphysical explanation (so-called ‘separatists’). I favour the latter view. See Raven (2015) for discussion.

(because its first premise is false), meaning that physicalism can be preserved even though we insist that Mary learns a genuinely new fact upon experiencing red.

#### 4 Subjective physicalism?

As we've interpreted the story, Mary does not know the phenomenal fact  $\Delta$  while still inside the room. Rather, she comes to know that fact only after leaving the room and having an experience of red. Notice, though, that this is not yet to insist that Mary *couldn't* have known  $\Delta$  whilst still inside the room. Indeed, it seems to me that there are various ways in which this might have happened. For instance, while still inside the room, Mary might have pushed on one of her closed eyelids, thus making red patches seem to appear; or else she might have induced in herself a red hallucinatory experience, perhaps by ingesting the right kind of substance, or even by directly manipulating her own brain. In one of these ways, Mary might well have experienced red even whilst still inside the black-and-white room, which would have enabled her to learn that  $\Delta$  obtains (cf. Dennett 2005; Johnston 2004).

The present view, therefore, does not commit us to the claim that Mary could not have known  $\Delta$  inside the room. However, one might think that it does commit us to a closely related thesis, namely, that Mary could not have known  $\Delta$  without undergoing an experience of red. Indeed, there are philosophers who have rejected premise 1 precisely by defending precisely this claim (see Alter 1998; Flanagan 1992; Howell 2007, 2008, 2009; cf. Crane 2003, 2019). On the view that these philosophers advocate, while phenomenal facts like  $\Delta$  are physical, they can be known *only* on the basis of specific kinds of conscious experience. Knowing  $\Delta$ , for instance, requires actually undergoing a conscious experience of red. And so the reason, on this view, that Mary did not know all of the physical facts whilst still inside the room is that by hypothesis, she had not yet experienced red, and hence had not met one of the necessary conditions on coming to know that  $\Delta$  obtains.<sup>23</sup>

Howell (2007, 2008, 2009) refers to the view as 'subjective physicalism'. The idea is that some facts, namely phenomenal facts, are both physical and subjective, in the sense of being knowable only to a certain kind of subjective point of view. More precisely, the claim is that knowing a phenomenal fact  $X$  requires having a specific conscious experience with the relevant phenomenal character  $C$ , so that in the absence of such experience, one couldn't know that  $X$  obtains.<sup>24</sup>

Underlying subjective physicalism is what we might call *the empiricist thesis*, to the effect that knowing a phenomenal fact  $X$  requires having a specific conscious experience, involving in some appropriate way the phenomenal properties  $X$  has among its constituents.<sup>25</sup> The trouble, however, is that there is reason to be suspicious of the empiricist thesis. Hence, one might think that if indeed the

---

<sup>23</sup> It is also a part of the phenomenal concept version of the "old fact/new guise" strategy for handling the knowledge argument that Mary could only learn what experiencing red is like once she (i) has the relevant phenomenal concept which (ii) could be acquired only on the basis of undergoing the relevant kind of experience. See, for example, Tye (1995).

<sup>24</sup> If  $X$  is a fact about a token experience  $e$  then this will be trivial: if the subject did not have that experience, then that experience could not exist nor have the character it does, and therefore  $X$  could not obtain. The factivity of knowledge would thus ensure that  $S$  knows  $X$  only if  $S$  experiences  $e$ . The empiricist thesis is thus non-trivial only as regards phenomenal facts about experience-types.

<sup>25</sup> I discuss this thesis (in a much more sympathetic vein) in Moran (manuscript).

grounding physicalist response to the knowledge argument requires commitment to subjective physicalism, this constitutes reason to be suspicious of that strategy.

Why be suspicious of the empiricist thesis? First, note there is a kind of case one can imagine that seems to put at least some pressure on physicalists to accept that having an experience of a certain type is not strictly necessary for coming to know what experiences of that type are like. Suppose, for example, that by some chance event – perhaps Mary slips and bangs her head on the wall – she ends up in precisely the same brain state as someone who (though not presently experiencing red) already knows what experiencing red is like. Indeed, by suitably manipulating the details of the case, we might suppose that Mary ends up as a complete microphysical duplicate of someone who knows that  $\Delta$  (but is not currently having a red experience). The thought is that given the not unreasonable idea that (at least within a physicalist framework) states of knowledge supervene upon the microphysical properties of subjects, it follows that in this variant case, Mary ends up knowing what red is like even without ever having experienced red. In turn, this would mean that the empiricist thesis is false. To know what red is like, Mary need not induce in herself an experience of red. All that she needs to do is put herself into the right kind of physical state (cf. Dennett: 2005).<sup>26</sup>

Second, one might well worry about the coherence of a view on which there are physical facts that can be known *only* on the basis of specific kinds of experience (cf. Crane 2003: 295). For, one might well think that physical facts are by their nature knowable from multiple different epistemic vantage points. Indeed, this is arguably one of the main lines of thought in Nagel (1986). According to Nagel, the physical world is also an *objective* world that is in principle knowable from many different perspectives. A constraint, therefore, on a fact being genuinely physical is that it be objective, which means at least that it cannot be knowable only from one specific vantage point, for instance only by creatures who are capable of having the precise kind of experience that normal perceivers have when sensing something red. If this idea is right, however, then it would be mistaken to maintain that some phenomenal facts are both (i) knowable only on the basis of undergoing specific experiences but also (ii) genuinely physical. Yet, this is precisely the view that those physicalists mentioned above (who reject premise 1 of the knowledge argument by endorsing the empiricist thesis) maintain. For, in their view, facts like  $\Delta$  are both phenomenal and physical, yet knowable only to subjects who are capable of undergoing sensory experiences with the requisite phenomenal characters.

Now it might turn out that the above two lines of thought can be resisted. Perhaps microphysical duplicates can differ with respect to knowledge of phenomenal properties.<sup>27</sup> And maybe some physical facts are indeed subjective in

---

<sup>26</sup> One might also take issue with the empiricist thesis itself even without presupposing physicalism. Take a theological example. If God is omniscient, then He could surely know what pain is like. But arguably, we do not want to say that a perfect being actually experiences pain. Thus a certain kind of theist might reasonably insist that God can come to know what pain is like without experiencing it; and then claim likewise for the other sensible qualities. (Cf. the discussion in Berkeley's *Third Dialogue* regarding God's knowledge of mental states such as being in pain.)

<sup>27</sup> A certain kind of phenomenal externalist, for example, might argue that phenomenal knowledge requires possession of the requisite phenomenal concepts, and that possession of such concepts depends on extrinsic causal-historical matters and can therefore differ across micro-duplicates. My point, however, is only that the above generates some reason for thinking that physicalists would be best not committing to the empiricist thesis.

the sense of being knowable only to specific experiential points of view. For present purposes, however, the thing to note is that there's no real need to settle this. The grounding physicalist need not accept the empiricist thesis (though equally she is quite free to adopt it if she wants). Rather, it is enough for her to insist that in the story as initially described, Mary *did not* know  $\Delta$ , and was moreover in a perfectly good sense not even in a position to know  $\Delta$ , due to not possessing the requisite phenomenal concept, namely PHENOMENAL RED. While still inside the room, Mary does not have this concept, and hence she cannot even entertain the proposition  $\Delta$  expresses. But once she leaves the room and experiences red for the first time, she gains the relevant concept, which makes it possible (or more exactly contributes to making it possible) for her to come to know that  $\Delta$  obtains. All of that, however, is quite consistent with also allowing that, at least in suitably unusual circumstances, Mary might somehow have obtained the concept PHENOMENAL RED without experiencing red. To be sure, the usual way of gaining that concept is by having an experience of the relevant kind. And it is plausibly the case that it's so difficult to come by the concept in a non-standard way that we can truly of Mary that she was in no real position to learn  $\Delta$  whilst still inside the room. But this is consistent with allowing that, in some sense, Mary *could* have learned  $\Delta$  while still inside the room and even without undergoing an experience of red. Accordingly, the physicalist can remain neutral about the empiricist thesis expressed above, thereby sidestepping potential objections along the lines of the two already canvassed.<sup>28,29</sup>

In light of the foregoing, perhaps the best way to express the present response to the knowledge argument would be as follows. If nothing unusual happens in the room, then upon leaving it and experiencing red for the first time, Mary will learn a new phenomenal fact (having gained, as a necessary precondition for learning this fact, the relevant phenomenal concept). The phenomenal fact she learns, however, will nonetheless be a physical fact. And so premise 1 of the knowledge argument will be false. However, if something unusual does happen in the room, such that in virtue of these unusual happenings, Mary gains the concept PHENOMENAL RED, then Mary might well learn what experiencing red is like even prior to escaping her achromatic chamber. Indeed, we might well grant that given a suitable array of unusual happenings, Mary might come to learn all of the physical facts there are

---

<sup>28</sup> It might well be that possessing a phenomenal concept  $C$  brings with it the capacity to imagine (or to remember or to q-remember [cf. Shoemaker 1970]) an experience of the associated phenomenal kind. In which case, a watered-down version of the radical empiricist thesis might seem to be preserved. In particular, what would remain is the idea that knowing e.g. what experiences of red are like involves the capacity to imagine (or recall or pseudo-recall, etc.) an experience of red.

<sup>29</sup> One might worry that something like Nagel's view about the necessary conditions for a fact to count as physical imperils my own view, on which  $\Delta$  is a derivative physical fact. After all, Nagel can plausibly be read as arguing for an epistemic constraint on physicality, such that a fact is physical only if it is knowable by the objective methods of the physical sciences. Yet even without commitment to the empiricist thesis, it seems that on my view, phenomenal facts like  $\Delta$  fail to meet this constraint. In response, I accept that facts like  $\Delta$  fail to meet this constraint. But I would offer the following modification to Nagel's constraint on physicality. Rather than saying that a fact is physical only if knowable by the objective methods of the physical sciences, we can instead offer a disjunctive account of physicality, such that a fact is physical only if either (i) it is a fundamental fact read off from our best physical science (and hence knowable by objective scientific means) or else (ii) metaphysically grounded therein. In this framework, the facts that meet condition (i) will be fundamental physical facts, whilst the facts that meet condition (ii) will be derivative physical facts.

(including all of the phenomenal facts) whilst inside the room.<sup>30</sup> In this alternate case, premise 1 will hold. However, it will no longer be plausible to think that Mary learns anything new when she leaves the room and experiences red. So in this variant case, premise 2 of the knowledge argument will be false. Either way, therefore, it emerges that the knowledge argument is not going to work.

## 5 Scrutability

Many philosophers in the mind-body literature are happy to grant a distinction between the narrowly physical facts (at the fundamental level) and the broadly physical facts (which are distinct from but dependent on the base physical facts). It is common, moreover, for such philosophers to maintain that the derivative, broadly physical facts are *scrutable* from their narrowly physical base. In other words, it is common to grant that - at least in ideal circumstances for rational reflection - one should be able to deduce a priori all of the non-fundamental facts from the sparse basis of fundamental facts on which they depend. This kind of thesis is often referred to as a scrutability thesis, the claim being that the derivative facts are scrutable from the fundamental ones. To focus on the case involving Mary, we end up with a specific thesis, which we can call *Scrutability*, implying that if Mary knew the full range of fundamental facts  $\Phi_1, \Phi_2, \Phi_n$ , she could deduce a priori all of the derivative facts, including the phenomenal fact  $\Delta$ . In our grounding physicalist framework, moreover, this amounts to the claim that if Mary knows the full range of fundamental facts then she should be able to deduce by a priori means the full range of further derivative facts that are metaphysically grounded therein.

I suspect that tacit commitment to *Scrutability* goes some way to explaining why so few philosophers in the literature have viewed rejecting premise 1 as a viable option for handling the knowledge argument. Indeed, commitment to *Scrutability* may even underlie Jackson's own acceptance of that premise, for which, again, he actually provides rather little argumentative support.<sup>31</sup> Moreover, *Scrutability*, if true, would certainly undermine the response to the knowledge argument we've been pursuing. To see this, note that that response entails that when Mary is still inside the black-and-white room, the following two claims jointly obtain, namely:

- (1) Mary knows  $\Phi_1, \Phi_2, \Phi_n$ .
- (2) Mary is not in a position to know  $\Delta$ .

---

<sup>30</sup> Well, not quite. Just as there might well be phenomenal states that creatures like Mary just cannot get into, so too, there might be phenomenal concepts that creatures like Mary just cannot come to possess. (This is especially likely if, as I speculated above, having a phenomenal concept necessarily goes along with being able to imagine an experience,  $e$ , a given token of which would instantiate the associated phenomenal property.) If this is correct, we get the interesting result that, contra the main text above, there are some phenomenal facts (and, hence, in the current framework at least, some genuinely physical facts) that are necessarily beyond Mary's ken, namely the phenomenal facts that are associated with the phenomenal concepts that she just cannot possess.

<sup>31</sup> Cf. especially Jackson (1986: 291), who just assumes at the outset that in the room Mary 'knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles'.

Again, even while still inside the room, Mary might come to learn  $\Delta$  if certain highly unusual happenings obtain, ones that enable her to acquire the concept PHENOMENAL RED in an abnormal way. But let us stipulate that no such unusual happenings obtain, and moreover that they could not have easily obtained. It follows that in the black-and-white room, Mary is not in a position to know  $\Delta$ . Given (1), however, if *Scrutability* holds, it follows that Mary *is* in a position to know  $\Delta$ , namely by means of a priori deduction. (Or at least, this follows given the further assumption, which I will grant in what follows, that in the room, Mary is in ideal circumstances for rational reflection. I see no reason not to accept this.)

Accordingly, it looks like we must reject *Scrutability*, i.e. deny that one can deduce a priori the full set of derivative facts from the array of fundamental facts. How troublesome is this? I will consider, and try to alleviate, four main concerns.

The first concern is that without *Scrutability*, the supervenience of the mental on the physical can't be accounted for. Of course, we've been operating with the notion of grounding rather than supervenience, but it is plausible to hold that grounding relations at least entail corresponding supervenience claims. This is because it is plausible to insist that grounds necessitate, so that if A grounds B, then necessarily, if A obtains, then B obtains.<sup>32</sup> With this in mind, consider again the grounding thesis (G) and the corresponding supervenience thesis (S):

$$(G) \quad \Delta < \Gamma_1, \Gamma_2, \dots \Gamma_n$$

$$(S) \quad \Box (\Gamma_1, \Gamma_2, \dots \Gamma_n \supset \Delta)$$

If grounds necessitate, then (G) of course entails (S). And since  $\Gamma_1, \Gamma_2, \dots \Gamma_n$  are properly among  $\Phi_1, \Phi_2, \dots \Phi_n$ , (S) in turn entails:

$$(S^*) \quad \Box (\Phi_1, \Phi_2, \dots \Phi_n \supset \Delta)$$

Again, however, many philosophers think that supervenience claims like (S\*) require a certain sort of backing. In particular, Jackson and Chalmers have argued that supervenience claims like (S\*) can be adequately explained or justified only by demonstrating that the A-facts comprising the supervenience base a priori entail the B-facts supervening on them.<sup>33</sup> If that is right, however, then (S\*) entails:

$$\textit{Scrutability}: \quad \Phi_1, \Phi_2, \dots \Phi_n \Rightarrow \Delta$$

(That is:  $\Delta$  is a priori entailed by the collection of facts:  $\Phi_1, \Phi_2, \dots \Phi_n$ .)

---

<sup>32</sup> That being said, the issues here get rather complicated rather quickly. For relevant discussion see Leuenberger (2014) and Moran (2021).

<sup>33</sup> Cf. Stoljar (2001: 254): 'Many philosophers hold that supervenience stands in need of justification or explanation; Jackson and Chalmers argue that the project of justifying or explaining supervenience just is the project of making it plausible that there is an a priori entailment of the mental by the physical'. Cf. Chalmers (1996, 1999), Chalmers & Jackson (2001); Jackson (1998).

This line of thought suggests that grounding physicalism implies *Scrutability*. But if that's true, then my proposed grounding physicalist response to the knowledge argument is undermined. I see at least two ways of pushing back.

First, there is room for grounding physicalists to deny that grounds necessitate, thus blocking the inference from (G) to (S). On the one hand, there is no consensus in the literature as to whether or not grounds necessitate: many philosophers accept this claim, but many others reject it. On the other hand, one might think that we have special reasons for thinking that while mental facts are grounded in physical facts, the physical facts in particular do not necessitate the mental facts they ground. (For instance, Moran 2021 argues that physicalists should reject necessitation in order to handle traditional conceivability/zombie arguments.) Thus, one way to block the above line of thought is to deny that grounded mental facts like  $\Delta$  are necessitated by the fundamental physical facts that ground them, either because no grounding facts necessitate the facts they ground, or because while some facts do necessitate the facts they ground, physical facts do not necessitate the mental facts they ground, this being a special feature of some grounding relations but not others, including physical to mental grounding relations.

Second, one could accept necessitation yet deny that supervenience claims have to be justified by underlying claims of a priori entailment. Indeed, one might reasonably think that grounding claims in and of themselves are sufficient to back the associated supervenience claims. As Bader (2017: 115) writes, 'grounding relations are precisely the kinds of explanatory relations that supervenience relations are meant to model and in terms of which supervenience claims can ultimately be explained and justified'. Accordingly, there's room to grant the inference from (G) to (S), and then from (S) to (S\*), but to resist the move from (S\*) to *Scrutability*.

A second line of argument for thinking that grounding physicalists must accept *Scrutability* is that since all of the other (non-mental) derivative facts *can* be derived a priori from the fundamental facts, we must accept that the same is true of the mental facts. In other words, the idea is that since we should accept claims like:

$$\textit{Scrutability-C} \quad \Phi_1, \Phi_2, \dots \Phi_n \Rightarrow C$$

$$\textit{Scrutability-B} \quad \Phi_1, \Phi_2, \dots \Phi_n \Rightarrow B$$

(where 'C' denotes the full range of chemical facts, and 'B' the full range of biological facts), there is pressure on the grounding physicalist to accept:

$$\textit{Scrutability-M}: \quad \Phi_1, \Phi_2, \dots \Phi_n \Rightarrow M$$

(where 'M' denotes the full range of mental facts), and therefore to accept:

$$\textit{Scrutability}: \quad \Phi_1, \Phi_2, \dots \Phi_n \Rightarrow \Delta$$

(given that  $\Delta$  is one of the mental facts). (Cf. Chalmers 1996). Again, however, there are two ways for the grounding physicalist to respond.

The first is to follow Schaffer (2017) in arguing that not even the chemical and biological facts (etc.) are a priori entailed by the fundamental physical facts.<sup>34</sup> Again, the line of argument above was that since (i) all non-mental higher-level facts are a priori entailed by the fundamental facts, (ii) it would be implausible to deny that the same is true of the mental facts. According to this first response, however, the first premise is mistaken: not all of the higher-level mental facts are a priori entailed by the fundamental physical facts. So, one does not end up implausibly claiming that the mental facts are special due to not being a priori entailed by their basic grounds.

Why think that not even the various non-mental higher-level facts, such as the chemical and the biological facts, are a priori entailed by the (basic) physical facts? Schaffer's argument draws on the contemporary debate about composition. Suppose we have three atoms: H, H and O, propertied and related in the kind of way that we ordinarily believe makes for the presence of an H<sub>2</sub>O molecule. We can distinguish two philosophical views. The nihilist maintains that there is no H<sub>2</sub>O molecule, only the three constituent atoms that we mistakenly believe, in everyday life, to compose a fourth and further complex thing. The non-nihilist, meanwhile, thinks that in addition to the atoms H, H and O there is a fourth and further thing that they compose, namely the H<sub>2</sub>O molecule. What Schaffer plausibly argues is that even if the non-nihilist is right, the nihilist's position is conceivable and consistent, and hence cannot be ruled out a priori even by an ideal mind. Generalising, the result is that even the chemical facts, such as the fact that there exists an H<sub>2</sub>O molecule, are not a priori derivable from the base facts.<sup>35</sup>

One can also apply similar reasoning to the biological case. A nihilist will say that there are atoms arranged organism-wise but no organism; a non-nihilist will say that in addition to the atoms arranged organism-wise there is an organism they compose. Again, the nihilist position seems coherent and consistent, so that not even an ideal mind could rule it out a priori. Accordingly, even if the non-nihilist is right (so that there are indeed biological facts), the biological facts do not appear to be a priori entailed by the underlying fundamental physical facts.

There is of course much more to say here. But I believe that the foregoing highlights the viability of the above line of response. Accordingly, I'll now turn to the second available response, namely denying premise (ii) of the argument, i.e. the claim that it would be implausible to insist that the mental facts (or maybe just a sub-set of the mental facts, such as the phenomenal facts) are special among higher-level facts insofar as they are not a priori derivable from the fundamental base.

Why would this be implausible? Why shouldn't the mental facts be special in this regard? What we need to avoid, of course, is accounting for the special status of the relevant mental facts in a way that jeopardises their claim to being physical. Fortunately, however, I believe that there is a plausible way of doing this.

The way to do so, I believe, is to again appeal to phenomenal concepts. Above, I suggested that in order to even entertain the proposition that the phenomenal fact  $\Delta$  expresses, Mary had to possess the relevant phenomenal concept. If that is right, however, then we can explain why, whilst still inside the room (when she

---

<sup>34</sup> For helpful discussion of Schaffer's argument cf. Aleksiev (2021); Rabin (2019); Sassarini (2021).

<sup>35</sup> Of course, the fact that the atoms H, H and O exist (and are propertied and related so) is plausibly not itself one of the fundamental facts; but we can use the above as a toy example.



does not possess that concept) Mary would be unable to deduce  $\Delta$  even despite knowing all of the relevant fundamental facts. The reason is that despite knowing the fundamental facts, since Mary lacks possession of the phenomenal concept needed to be able to entertain the proposition  $\Delta$  expresses, she is unable to come to know that  $\Delta$  and hence unable to deduce  $\Delta$  from what she already knows.<sup>36</sup>

Suppose now that Mary acquired the concept PHENOMENAL RED. Would she then be in a position to deduce  $\Delta$  from the fundamental facts  $\Phi_1, \Phi_2, \dots, \Phi_n$ ? In fact, I think that even this is questionable, and that the answer will depend on what we say about the status of the relevant bridge principles or grounding laws. In my preferred framework, analogously to causation, grounding operates in accordance with (governing) metaphysical laws whose role is to specify what grounds what. Hence, merely knowing the fundamental physical facts will be insufficient for being able to deduce the phenomenal facts, even given possession of the concept PHENOMENAL RED. For, in addition to knowing the fundamental physical facts, one would also need to know the metaphysical laws specifying that given such and such fundamental facts as input one gets certain derivative phenomenal facts as output.

Thus, if all that Mary knows inside the room are the fundamental facts, and if she neither knows the laws nor has the relevant phenomenal concept, she would be unable to deduce  $\Delta$ .<sup>37</sup> To be in a position to make that deduction, she would require not only the concept PHENOMENAL RED, but also knowledge of the relevant grounding law(s) specifying that with  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  as input one gets  $\Delta$  as output.<sup>38</sup>

The third concern draws on the fact that grounding claims are said to underwrite metaphysical explanations. For instance, if the plurality of fundamental facts  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  ground the phenomenal fact  $\Delta$ , they also metaphysically explain why  $\Delta$  obtains. One might argue, however, that a constraint on adequate metaphysical explanation is that the explanandum should be a priori deducible from the explanans. After all, if  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  explain  $\Delta$ , then they *make intelligible* why  $\Delta$  obtains. But how could it be that  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  be make intelligible why  $\Delta$  obtains unless it were also the case that once one knows that  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  obtain, one is thereby in a position to just see, and so deduce, that  $\Delta$  obtains as well?<sup>39</sup>

---

<sup>36</sup> In a variation on the view expressed above, Montero (2007: §IV) argues that Mary *can* deduce the phenomenal facts a priori from the fundamental physical facts even though she requires experience to gain the relevant phenomenal concept and even though possessing the relevant concept is a condition on being able to make the relevant deduction. Crucial to making out this view is the idea that a priori deductions are not deductions carried out in the absence of experience but rather deductions that can be *justified* independently of experience (i.e. ‘from the armchair’).

<sup>37</sup> Unless, of course, the metaphysical laws themselves can be derived a priori from knowledge of the fundamental grounding facts. I’m tempted to deny this, but I acknowledge that this requires considerable further discussion.

<sup>38</sup> Cf. Schaffer (2017: 18) ‘I think that [in the room] Mary suffers from a twofold deprivation. First, she has been deprived of metaphysical information insofar as she has not learned the relevant bridge principles. Secondly, she has been deprived of conceptual information, insofar as — barring additions to the story — she does not have the full phenomenal concepts needed to know what it is like...to experience a rose or a sunset’. (For present purposes, we can leave open the important question whether Mary could come to learn the relevant grounding laws a priori.)

<sup>39</sup> Compare here the literature on the ‘explanatory gap’. At least on one widely held construal, for there to be an explanatory gap between the mental and the physical is precisely for the mental facts to fail to be a priori deducible from the physical facts. But this clearly assumes that there’s a

Once again two things can be said in response. On the one hand, by stressing the widely remarked upon close analogy between causation and grounding, one can plausibly insist that a priori entailment is not necessary for metaphysical explanation. Suppose we ask why the radiator is hot.<sup>40</sup> A perfectly good causal explanation could be given by answering that someone turned the heating on at time *t*. Nor is this any way undermined by the fact that (as Hume taught us) one cannot deduce a priori the occurrence of a given effect from knowledge of its cause. In asking why the radiator is hot, however, we might instead be seeking a metaphysical (or constitutive) explanation, as opposed to a causal one. And here the answer that cites the motion of the molecules composing the radiator would be entirely sufficient. Again, however, it seems at least not obvious that this explanation would be way undermined if one denied that the derivative fact (that the radiator is hot) is a priori deducible from the underlying base facts.<sup>41</sup>

Secondly, and not incompatibly with the first response, one could grant that there is a necessary connection between metaphysical explanation and a priori entailment, but insist that the entailment is not merely between the grounding and the grounded facts, but rather between the grounding facts taken together with the metaphysical laws. More precisely, one could insist that grounding backs metaphysical explanation in the sense that if one knows *both* the grounding facts *and* the relevant bridge principles (and if one has the requisite conceptual repertoire) then this will make intelligible exactly why the grounded fact obtains, and so give one the capacity to deduce the grounded fact from those background conditions. This response preserves something of the rationalist idea that metaphysical explanation goes along with intelligibility and a priori deducibility, but is also quite compatible with the interpretation of the Mary story we have given.

The fourth and final concern is similar in certain ways to the third concern we just considered. The worry turns on the idea is that in general, it is possible to deduce which grounding claims obtain from knowledge of the relevant grounding (base) facts. This implies, quite generally, that if *X* grounds *Y*, then the knowledge that *X* obtains is sufficient for being able to deduce that *X* grounds *Y* and hence that *Y* obtains. The principle can be supported by examples. Consider these:

- (1) *x* is red or brown < *x* is red
- (2) *x* is red < *x* is scarlet
- (3) there is a party going on < people are reveling

In all these cases, it seems, it is plausible to think that knowledge of the fact doing the grounding (e.g. *x* is red) is sufficient for knowledge that the grounding relation in question (that *x* is red or brown in virtue of being red) obtains, and hence for

---

clear a link between failure of a priori deducibility (from the physical to mental) on the one hand and lack of adequate metaphysical explanation (of the mental in terms of the physical) on the other.

<sup>40</sup> The following example is adapted from Dasgupta (2017).

<sup>41</sup> Of course, one might insist that in the case of heat facts and molecular motion facts the relevant a priori entailment *does* hold. If so, I suggest that the reader replace this example with one this is less obviously the case, e.g. the claim that an action's being wrong is metaphysically explained (though not all obviously a priori entailed by) its involving the promotion of general well-being.

knowledge that the grounded fact obtains (that  $x$  is red or brown) obtains. It seems, moreover, that one could easily come up with a range of further examples.

If, however, one can deduce which grounding relations hold just from knowledge of the relevant base facts, then if Mary knows that  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  she should be able to deduce that  $\Delta < \Gamma_1, \Gamma_2, \dots, \Gamma_n$  and hence that  $\Delta$ . Therefore, since  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$  are properly amongst  $\Phi_1, \Phi_2, \dots, \Phi_n$ , it follows that knowledge that  $\Phi_1, \Phi_2, \dots, \Phi_n$  obtain should be sufficient for deducing  $\Delta$ , as per *Scrutability*.

The way to respond here, I think, is to deny that in general, if  $X$  grounds  $Y$ , then knowing that  $X$  obtains is sufficient for being able to deduce that  $X$  grounds  $Y$  and hence that  $Y$  obtains. Here are some relevant comments about the above examples. With (1) and (2), the lines of reasoning involved plausibly rely on grounding principles that can indeed be known a priori, namely a principle concerning disjunction and a principle concerning determinables and determinates respectively.<sup>42</sup> Were these principles not knowable a priori, however, it would no longer be plausible to maintain that (1) and (2) can be deduced from the fact doing the grounding alone. As for (3), the crucial point seems to be that the *metaphysical analysis* or *real definition* of the grounded fact can be given in terms of the fact doing the grounding, and moreover one can know this definition or analysis a priori. That is, it is knowable a priori that *what it is* for there to be a party going on is for there to be some people (collectively) revelling.<sup>43</sup> Hence, one can know a priori that if some people are, then there is for that reason a party going on.

It is unclear, however, that all grounding claims have to be like this. Indeed, in the case of physical to mental grounding, I'd be tempted to deny that either situation holds. First, I don't think that one can give a real definition of the phenomenal in terms of the fundamental physical at all, and so no such definition is knowable either a priori or otherwise. Second, I doubt that the grounding principles in accordance with which the fundamental physical gives rise to the phenomenal are knowable only on the basis of knowing the fundamental physical facts themselves.<sup>44</sup> Nor do I think the mental is a special case here. Rather, I think that much the same can be said for the way in which the normative is grounded in the descriptive, or the way in which the sensible qualities of objects are grounded in their non-sensible physical properties (cf. Moran 2023). Such cases of grounding strike me as perfectly intelligible. Accordingly, I think there's good reason to deny that in general, if  $X$  grounds  $Y$ , then the knowledge that  $X$  obtains is sufficient for being able to deduce that  $X$  grounds  $Y$  and hence that  $Y$  obtains.<sup>45</sup>

In sum, then, it turns out that we can coherently reject *Scrutability*, and thus avoid the worry we began with in this section. In this way, we can retain the view

---

<sup>42</sup> For helpful discussion of the relevant grounding principles here see Rosen (2010).

<sup>43</sup> As Selim Berker pointed out to me, if some people were individually revelling in the same location, perhaps for independent reasons, it is unclear that their activities would constitute a party.

<sup>44</sup> Does this imperil the claim, made earlier in §3, that since the mental is grounded in the physical, grounded mental facts in some sense consist in the physical facts that ground them? I think not, but the issue is a complicated one I lack space to discuss here (but see Moran 2023).

<sup>45</sup> It might nevertheless be so that grounding facts can be read off knowledge of the grounded facts. As Fine (2012) points out, it is plausible to think in general that it is grounded facts that 'point to' their (actual and possible) grounds, rather than things being the other way around.

that where the knowledge argument goes wrong is with the idea that even in the room prior to experiencing red Mary knows all of the available physical facts.<sup>46</sup>

## 6 Conclusion

The knowledge argument turns on two main premises:

- 1 In her black and white room, Mary knows all the physical facts.

And

- 2 When Mary experiences red for the first time, she learns a new fact.

Again, it is widely assumed that physicalists must respond to the argument by rejecting the second premise and granting the first. In this paper, however, I have shown that by operating in a grounding physicalist framework, physicalists can instead reject the first premise and accept the second. This then enables them to respect the very plausible idea that upon leaving the room and experiencing red for the first time, Mary learns a genuinely new phenomenal fact, concerning what experiences of red are like. Traditionally, this intuition has been viewed as being the sole preserve of non-physicalist views, such as dualism or panpsychism. What I have shown here, however, is that by adopting a grounding framework, it emerges that physicalists are just as well placed as their opponents to accept this claim.

## References

- ALEKSIEV, D. (2022) 'Lightweight and Heavyweight Anti-Physicalism', *Synthese*, 200 (112): 1-23.
- ALTER, T. (1998) 'A Limited Defense of the Knowledge Argument', *Philosophical Studies*, 90 (1): 35-56.
- AUDI, P. (2012) 'A Clarification and Defense of the Notion of Grounding', in Correia, F. & B. Schnieder (eds.) *Metaphysical Grounding: Understanding the Structure of Reality*, pp. 102—121, Cambridge: Cambridge University Press.
- ARMSTRONG, D. M. (1997) *A World of States of Affairs*, (Cambridge: Cambridge University Press).
- BADER, R. M. (2017) 'The Grounding Argument Against Non-Reductive Moral Realism', *Oxford Studies in Metaphysics* 12, (Oxford: Oxford University Press).

---

<sup>46</sup> Of course, there might be plausible alternate scrutability theses in the vicinity that we might still accept, e.g. that anyone who knows the base facts, has the requisite concepts, and knows the requisite grounding laws will be able to deduce the derivative facts. Cf. Schaffer (2017: 18-19).

- (forthcoming) ‘Grounding, Reduction and Analysis’, to appear in *Objects and Properties*, (eds) Moran, A. & Rossi, C. (Oxford: Oxford University Press).
- BENNETT, J. (2001) *Learning From Six Philosophers: Descartes, Spinoza, Leibniz, Berkeley, Locke, Hume*, Vol. 2, (Oxford: Clarendon Press).
- BROAD, C. D. (1925), *The Mind and its Place in Nature*, (London: Routledge).
- CHALMERS, D. (1996) *The Conscious Mind* (New York: Oxford University Press).
- (1999) ‘Materialism and the Metaphysics of Modality’, *Philosophy and Phenomenological Research*, 59 (2): 473-496.
- CHALMERS, D. & JACKSON, F. (2001) ‘Conceptual Analysis and Reductive Explanation’, *Philosophical Review*, 110 (3): 315-61.
- CONNOR, E. (1994) ‘Phenomenal Knowledge’, *Australasian Journal of Philosophy*, 72: 136-150.
- CRANE, T. (2003) ‘Subjective Facts’, in *Real Metaphysics*, (eds.) Lillehammer, H. & Rodriguez Pereyra, G. (eds.), (London: Routledge): 68-83.
- (2019) ‘The Knowledge Argument is an Argument About Knowledge’, in *The Knowledge Argument*, (ed.) Coleman, S., (Cambridge: Cambridge University Press): 15-31.
- DASGUPTA, S. (2014) ‘The Possibility of Physicalism’, *Journal of Philosophy*, 111 (9- 10): 557-592.
- (2017) ‘Constitutive Explanation’, *Philosophical Issues*, 27 (Metaphysics): 74-97.
- DENNETT, D. (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, (Harvard: MIT Press).
- DEROSSET, L. (forthcoming) *Fundamental Things*, (Oxford: Oxford University Press).
- DORSEY, J. E. (2016) ‘On the Ground-Reduction Link’, *American Philosophical Quarterly* 53 (4): 411-422.
- FINE, K. (1982) ‘First-Order Modal Theories III: Facts’, *Synthese*, 53 (1): 43-122.
- (2012) ‘Guide to Ground’, in Correia, C. & Schnieder, B. & (eds.) *Metaphysical Grounding: Understanding the Structure of Reality*, (Cambridge: Cambridge University Press): 37-80.

- FLANAGAN, O. (1992) *Consciousness Reconsidered*, (Cambridge MA: M.I.T. Press).
- FREGE, G. (1918–1919) ‘Thoughts’ in N. Salmon and S. Soames (eds.) *Propositions and Attitudes* (Oxford: Oxford University Press 1988): 33–55.
- GOFF, P. (2017) *Consciousness and Fundamental Reality*, (Oxford: Oxford University Press).
- GREGORY, D., HENDRICKX, M. & TURNER, C. (2022) ‘Who Knows What Mary Knew? An Experimental Study’, *Philosophical Psychology*, 35 (4): 522–545.
- HATTIANGADI, A. (2018) ‘Moral Supervenience’, *Canadian Journal of Philosophy*, 48 (3/4): 592–615.
- HOWELL, R. (2007) ‘The Knowledge Argument and Objectivity’, *Philosophical Studies*, 135: 145–177.
- (2008) *Consciousness and the Limits of Objectivity*, (Oxford: Oxford University Press).
- (2009) ‘The Ontology of Subjective Physicalism’, *Noûs*, 43 (2): 315–345.
- HORGAN, T. (1984) ‘Jackson on Physical Information and Qualia’, *Philosophical Quarterly*, 32: 127–136.
- (1993) ‘From Supervenience to Superdupervenience: Meeting the Demands of a Material World’, *Mind*, 1993, 102: 555–86.
- JACKSON, F. (1982) ‘Epiphenomenal Qualia’, *The Philosophical Quarterly*, 32 (127): 127–136.
- (1986) ‘What Mary Didn’t Know’, *The Journal of Philosophy*, 83 (5): 291–295.
- (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. (Oxford: Clarendon Press).
- (2003) ‘Mind and Illusion’, *Royal Institute of Philosophy Supplement*, 53: 251–271.
- JOHNSTON, M. (2004) ‘The Obscure Object of Hallucination’, *Philosophical Studies*, 120 (1/3): 113–183.
- (2006) ‘Hylomorphism’, *Journal of Philosophy*, 103 (12): 652–698.
- KIM, J. (1987) ‘“Strong” and “Global” Supervenience Revisited’, *Philosophy and Phenomenological Research*, 48, 2: 315–326.

- (1998) *Mind in a Physical World*, (Cambridge, MA: MIT Press).
- (2002) ‘The Layered Model: Metaphysical Considerations’, *Philosophical Explorations*, 5 (1): 2-20.
- LEUENBERGER, S. (2014) ‘From Grounding to Supervenience?’, *Erkenntnis*, 79 (1): 227-240
- LEVIN, J. (2007) ‘What is a Phenomenal Concept?’, in *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, (eds.) Alter, T. & Walter, S. (Oxford: Oxford University Press): 87-110.
- LEWIS, D. K. (1988) ‘What Experience Teaches’, *Proceedings of the Russellian Society*, 13: 29-57
- (1994) ‘Humean Supervenience Debugged’, *Mind*, 103 (412): 473-490.
- LIU, M. (forthcoming) ‘Revelation and the Appearance/Reality Distinction’, to appear in *Oxford Studies in Philosophy of Mind*, vol. 4.
- LOAR, B. (1990), ‘Phenomenal States’, *Phenomenal Perspectives 4: Action Theory and Philosophy of Mind*: 8-108.
- MOORE, G. E. (1903) *Principia Ethica*, (Cambridge: Cambridge University Press).
- MONTERO, B. (2007) ‘Physicalism Could Be True Even If Mary Learns Something New’, *Philosophical Quarterly*, 57 (227): 176-189.
- MORAN, A. (2021) ‘Grounding the Qualitative: A New Challenge for Panpsychism’, *Journal of Consciousness Studies*, Special Issue, Goff, P & Moran, A. (eds), 28 (9/10): 163-180.
- (2022) ‘Living Without Microphysical Supervenience’, *Philosophical Studies*, 179 (1): 405-428.
- (2023) ‘Moorean Connections and Grounding Laws’, *Inquiry*, [doi.org/10.1080/0020174X.2023.2253863](https://doi.org/10.1080/0020174X.2023.2253863)
- (manuscript) ‘Primitive Colours in a Physical World’, m.s. draft.
- NAGEL, T. (1986) *The View From Nowhere*, (Oxford: Oxford University Press).
- NEMIROW, L. (1980) Review of Thomas Nagel, *Mortal Questions*,’ *Philosophical Review*, 89: 473-477.

- NIDA-RÜMELIN, M. (2007) 'Grasping Phenomenal Properties', in *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, (eds.) Alter, T. & Walter, S. (Oxford: Oxford University Press).
- NIDA-RÜMELIN, M. & O'CONNELL, D. (2021) 'Qualia: The Knowledge Argument', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2021/entries/qualia-knowledge/>>.
- (forthcoming) 'Is Dualism Compatible with Consciousness Being Grounded in the Physical?', to appear in *Grounding and Consciousness*, (ed.) Rabin, G. O., (Oxford: Oxford University Press).
- OPPENHEIM, P. & PUTNAM, H. (1958). 'Unity of Science as a Working Hypothesis', in *Minnesota Studies in the Philosophy of Science*, Volume II, (eds.) Feigl, H., Scriven, M. & Maxwell, G., (University of Minnesota Press): 3-36.
- PAUTZ, A. (forthcoming) 'How to Achieve the Physicalist Dream Theory of Consciousness: Identity or Grounding?', to appear in *Grounding and Consciousness*, (ed.) Rabin, G. O., (Oxford: Oxford University Press).
- PERRY, J. (2001) *Knowledge, Possibility, and Consciousness*, (Cambridge MA: MIT Press).
- RABIN, G. O. (2019) 'Grounding the Gaps or Bumping the Rug? On Explanatory Gaps and Metaphysical Methodology', *Journal of Consciousness Studies*, 26 (5/6): 191-203.
- RAVEN, M. (2015) 'Ground', *Philosophy Compass*, 10 (5): 322-333.
- ROBINSON, H. (1982) *Matter and Sense: A Critique of Contemporary Materialism*, (Cambridge: Cambridge University Press).
- (2016) *From the Knowledge Argument to Mental Substance: Resurrecting the Mind*, (Cambridge: Cambridge University Press).
- ROSEN, G. (2010) 'Metaphysical Dependence: Grounding and Reduction', in Hale, B & Hoffmann, A. (eds.) *Modality: Metaphysics, Logic, and Epistemology*, (Oxford: Oxford University Press): 109-136.
- RUSSELL, B. (1927) *The Analysis of Matter*, (London: Kegan Paul).
- SASSARINI, E. (2021) 'No Ground to Bridge the Gap', *Synthese*, 199: 7981-7999.
- SCHAFFER, J. (2009) 'On What Grounds What', in Chalmers, D, Manley, D. & Wasserman, R. (eds.) *Metametaphysics*, (Oxford: Oxford University Press) 357-383.



- (2010) ‘Monism: The Priority of the Whole’, *Philosophical Review*, 119 (1): 31-76.
- (2017) ‘The Ground Between the Gaps’, *Philosophers' Imprint* 17: 1-26
- (manuscript) ‘Naturalistic Dualism and the Problem of the Physical Correlate’, manuscript draft.
- SHOEMAKER, S. (1970) ‘Persons and Their Pasts’, *American Philosophical Quarterly*, 7 (4): 269-285.
- STOLJAR, D. (2001) ‘Two Conceptions of the Physical’, *Philosophy and Phenomenological Research*, 62 (2): 253-281
- (2005) ‘Physicalism and Phenomenal Concepts’, *Mind and Language*, 20 (5): 469-494.
- TROGDON, K. (2017) ‘Revelation and Physicalism’, *Synthese*, 194 (7): 2345-2366.
- TYE, M. (1995) *Ten Problems of Consciousness*, (Cambridge, MA: MIT Press).
- (2009) *Consciousness Revisited: Materialism without Phenomenal Concepts*, (Cambridge, MA: MIT Press).
- WITTGENSTEIN, L. (1922) *Tratatus Logico-Philosophicus*, (London: Kegan Paul).