I don’t believe that the considerations of conceivability discussed in the last chapter get to the heart of the anti-physicalist intuition in relation to consciousness. The chain of reasoning from (a) the conceivability of certain scenarios, to (b) the possibility of those scenarios, to (c) the actual falsity of physicalism, is a fairly fancy form of argumentation. There is a more immediate and widespread conviction that that – a feeling of pain – is not the same thing as that – a brain state. If this conviction is rational, rather than simply something we are inclined to believe for psychological reasons, then it must result from an understanding of the nature of those properties. It must be that (a) we grasp what it is for someone to feel pain, and (b) we grasp what it is (at least in general) for someone to instantiate a certain causal-structural state, and hence (c) it is apparent to our understanding that these are not the same thing.
This is indeed what I think is going on at the heart of the conviction that feelings can’t possibly be brain states. In this chapter I will defend this conviction, and articulate it as an argument against physicalism.¹

As in the last two chapters, my immediate target is pure physicalism: physicalism in conjunction with the thesis that fundamental reality can be exhaustively described in the mathematico-nomic vocabulary of physics. In chapter 6 I will extend the argument such that it targets physicalism in its general form.

Revelation and transparency

I believe that we are in a unique epistemic situation with respect to our conscious states. A normal human being is able to attend to one of the conscious states she is currently instantiating, and form a direct phenomenal concept of that state: a phenomenal concept of a conscious state the content of which is wholly based on attending to that state.² In having a direct phenomenal concept, the token conscious state being attended to is directly presented to the concept user in such a way that (i) the complete nature of the type to which it belongs is apparent to the concept user, and (ii) the concept user knows with certainty (or something close to it) that the token conscious state exists. To take a concrete example, when I attend to a specific pain and form a direct phenomenal concept of it, that token pain is directly presented to me; and because it is ‘right there’ for me (i) I know exactly what it is for someone to feel that way, and (ii) I know with certainty (or something very close to it) that I am in pain. Call this hypothesis ‘Revelation.’³

¹ I proposed versions of this argument in Goff 2011 and 2015b. See also Nida-Rümelin 2007, and the argument outlined, although not endorsed, in Lewis 1995.
² The notion of a direct phenomenal concept comes from Chalmers 2003.
³ The word ‘revelation’ was first used by Mark Johnston (1992) to describe Galen Strawson’s (1989) view that the nature of colour is wholly given in colour experience. The term later became used in the mind-body debate to refer to the view that the nature of experience is given in experience, e.g. Stoljar 2006. Note that I am using it here in a slightly more specific sense.
We should note that Revelation is not the same as Phenomenal Transparency, although the former entails a qualified form of the latter. Phenomenal Transparency is the thesis that phenomenal concepts reveal the essence of the states they denote. According to Revelation, when a person attends to a token conscious state in a direct phenomenal concept the complete nature of the type to which it belongs state is apparent to her; this entails direct phenomenal transparency: the thesis that direct phenomenal concepts are transparent. However, the entailment does not go the other way. Revelation involves the nature of the property instance being directly presented to one; it is because the concept application involves direct presentation that the nature of the property type is known (and the property token known with certainty to exist). But one could hold that phenomenal concepts are transparent, whilst denying that they are transparent because applications of the concept involve a direct presentation of a token of the referent. Presumably analytic functionalists have this view: conscious states are causal/functional properties, and paradigmatic concepts of them pick them out as such. More generally, logical and mathematical concepts are transparent, and yet applications of such concepts do not involve direct presentations of logical and mathematical properties (or at least not obviously so).

Some will be content to take Revelation as a bedrock assumption, a thesis which is evident enough not to need independent support. This does not mean taking the thesis itself to be certain; it just means that it is to be believed, or at least that it is reasonable to believe it, in the absence of strong evidence or argument to the contrary. I would like to invite the reader to spend a few minutes reflecting on the thesis for herself, and deciding whether it seems to her self-evident. William James declared that metaphysics is nothing other than an unusually obstinate attempt to think clearly. I think that’s an exaggeration, but it’s perhaps underrated in contemporary analytic metaphysics how powerfully persuasive calm and

---

4 Perhaps we should interpret analytic functionalists as not believing in phenomenal concepts. But if there were an analytic functionalist who wholeheartedly accepted the existence of phenomenal concepts, then it seems that such a person would take phenomenal concepts to be transparent.

5 James 1890/1981.
careful meditation on a doctrine can be.\textsuperscript{6} I offer the following (clearly biased) presentation of the thesis in question as a focus for the reader’s calm and careful meditation:

Surely, you know exactly what your pain is – what it is for someone to feel pained in precisely that way – just by attending to pain and thinking about in terms of how it feels. There is nothing in any way hidden from you about the reality of how you’re feeling; nor is it possible that you’re not really feeling that way. And that’s because the feeling is ‘right there’ for you in such a way that its reality cannot be doubted.

Please don’t read on until you have taken a few minutes to consider this.

If as the result of this meditation you are content to take Revelation as bedrock, premise 3 of the Transparency Conceivability Argument is already secure, so long as we take ‘Q’ to express a direct phenomenal concept (although we are yet to consider the truth of premise 2). For Revelation entails Direct Phenomenal Transparency, the thesis that direct phenomenal concepts are transparent.

\textbf{Transparency Conceivability Argument}

\textit{Premise 1} – ‘P and nothing has Q’ is conceivably true.

\textit{Premise 2} – (TCP) If a transparent sentence is conceivably true, then it’s possibly true.

\textit{Premise 3} – ‘P and nothing has Q’ is transparent.

\textit{Conclusion} – ‘P and nothing has Q’ is possibly true, and so pure physicalism is false.

Readers who don’t need to be further persuaded of the truth of Revelation may skip to the section \textit{The Conceivability Argument and the Transparency Argument}. However, I suspect that many readers did not come out of meditation persuaded of the evident truth of

\textsuperscript{6} I have often found contemporary metaphysicians giving the instantaneous reply ‘I don’t have that intuition’, as though consulting one’s intuitions took no longer that checking one’s iPhone.
Revelation. Therefore, in what follows I will try to build a case for it in terms of more obvious premises.

**The case for Revelation**

I will begin building the case for Revelation with the observation that introspection reveals substantive knowledge concerning our conscious states. Perhaps the most straightforward examples are within colour experience: I can know through introspection that what it’s like to see red is similar to what it’s like to see orange. It is also plausible that we know through introspective reflection that pain is ceteris paribus a bad thing: knowing what it’s like to feel pain reveals to me what a terrible thing it is to be in pain. Another plausible example: attention to my conscious states make me aware of their representational properties, for example, attention to my current visual experience of a tomato reveals it to be an experience as of a red, round thing at a certain distance from me. In order to avoid getting into more controversy that is necessary, I’ll focus on introspective knowledge of the resemblances between our sensory states, which is perhaps the most uncontroversial case of this kind of knowledge.

The mere fact that we have substantive information through introspection does not entail Revelation. The physicalist will no doubt simply postulate a faculty of introspection, analogous to our perceptual faculties; a kind of ‘inner sense’ through which we learn about our consciousness. The strange thing about introspection, though, is that it yields beliefs which are as certain as our beliefs concerning mathematics and logic.

Suppose you are in an exhibition of contemporary abstract art, staring at work called ‘Red and Orange Splodges.’ The work consists of a huge, homogenous patch of orange paint – which fills most of the left half of your visual filed, and a huge homogenous patch of red paint – which fills most of the right half of your visual field. Contrast the following two judgments:

- The *perceptual judgment* that there are in front of you – out there in the mind-independent world – two similar splodges of paint.
The introspective judgment, based on attending to your experience, that you are currently having two similar experiences

It is very easy to entertain sceptical doubts about the perceptual judgment. Perhaps I am having a particularly vivid – although admittedly rather dull – dream of being in an art gallery looking at splodges of paint, when in fact I am tucked up in bed with my eyes closed. Perhaps I am in the Matrix, being made to experience a virtual art gallery, when in fact the evil computers have long since destroyed real world galleries. The introspective judgement, in contrast, is much harder to doubt. Even supposing I am asleep or in the Matrix, it’s hard to persuade myself that what it’s like for me to experience that – the left-hand splodge – is similar to what it’s like for me to experience that – the right-hand splodge.

One might take it that the above merely reflects something psychological about human beings, or perhaps about human beings educated in a certain philosophical tradition. On this view we just happen to find it very hard to doubt certain of our introspective judgements; as perhaps we also find it hard to believe that we don’t have libertarian free will, or that the universe has no purpose. However, I want to suggest that the difference outlined above reflects not merely a psychological fact about us, but a deep and important fact about our epistemic situation. It is the foundational Cartesian truth, and it stands as good today as it did when Descartes first declared it: facts about one’s own experience are known with much greater certainty than facts about the external world.⁷

What is certainty? It is important to distinguish between the psychological fact that we have a certain amount of confidence in a given proposition, and the epistemic fact concerning the degree of confidence one is entitled to feel in that proposition (taking into account, for example, ones evidence). A person might be wildly confident that climate change is caused by the moon, whilst in fact – because she lacks any evidence – that person ought to have little or no confidence in that proposition. We can say that such a person has a high degree of ‘psychological confidence’, but that that psychological confidence is not ‘rational.

⁷ Augustine (426/1998) had previously defended a similar argument. See Ben-Yami 2015 for an interesting discussion of the two arguments.
confidence’. When a person has a rationally permissible degree of confidence in a given proposition, we can say that that person has ‘rational confidence’ in that proposition.

We are entitled to be much more confident of basic truths of maths and logic, than we are of the basic truths of perception. That is to say, I can achieve much higher levels of rational confidence that $1+1=2$ than I am that there is a table in front of me. This is reflected in the fact that it is much easier to entertain sceptical doubt about the latter fact than the former. I submit that many of our introspective judgements about our conscious experience are such that the degree of confidence we are entitled to hold in them is roughly the same as the degree of confidence we are entitled to hold in our basic mathematical judgements. I am entitled to be as confident that *this experience* – the experience of an orange patch – is similar to *that experience* – the experience of a red splodge – as I am that $2+2=4$. I will express this claim by saying that certain introspective judgments are ‘super-justified’.

I am not saying that either mathematical or introspective beliefs are *rationally certain*, i.e. such that we are entitled to believe them with a credence of 1. Perhaps all beliefs are such that we ought to entertain *some* degree of doubt; perhaps the evil demon is making it seem evident that $2+2=4$, when in fact $2+2=5$. And of course within the class of mathematical beliefs, and within the class of introspective beliefs, the degree of permissible confidence will vary; complex calculations, or introspective judgements made without concentrating, will incur a much greater possibility of error. This is all consistent with the claim that in general the degree of confidence it is permissible to have with regards to perceptual beliefs is much higher than the degree of confidence it is permissible to have with regards to mathematical and introspective beliefs, and that the degree of confidence it is permissible to have with regards to the latter two classes of beliefs is roughly similar. We can call this thesis ‘Super-Justification.’

There will of course be philosophers who will deny Super-Justification, insisting that what I have described above are merely psychological facts about human beings, or perhaps human beings infected by a Cartesian philosophical heritage. It could be claimed that, although mathematical truths are known with much greater certainty than empirical truths, empirical and introspective truths are in the same epistemic boat. Or it could even be denied that mathematical truths are known with greater certainty than empirical beliefs.
Perhaps the fact that we find it much harder to entertain sceptical doubts concerning $2+2=4$ simply reflects a psychological fact about us.

At this point we have reached bedrock. The thesis that introspective truths are super-justified – in the sense that they are known with a degree of rational confidence roughly similar to that of mathematical and logical beliefs, and that this is significantly higher than the rational confidence associated with perceptual beliefs – seems to me as evident as any philosophical starting point. Super-Justification is the rock upon which I shall build my argument.

Although it seems evident, if it is true that introspective judgements are super-justified, this is a startling fact about the human situation. It is one thing to know with great certainty facts about the Platonic realm – that $2+2=4$, or that square circles are impossible – but how is it that I can know with a similar degree of certainty a fact about the concrete, contingent world, e.g. that a particular experience I am currently having resembles some other particular experience I am currently having?

I submit that the best explanation of super-justification is Revelation. Return to the experience of red and orange splodges. According to Revelation, when I attend to the token conscious state representing the red splodge, that state is directly presented to me in such a way that (i) the complete nature of the type of which it is a token is directly presented to me in such a way that its reality cannot be rationally doubted, and (ii) in forming a direct phenomenal concept of the experience of the orange splodge its complete nature is directly presented to me in such a way that its reality is known with great rational confidence. In grasping the nature of each phenomenal property, I grasp that they have similar natures. And given that I have great rational confidence in the reality of each experience, having the nature each does, I thereby know with a great degree of certainty that they resemble. Revelation explains Super-Justification.

This explanation might seem somewhat ad hoc and unilluminating. Of course if you postulate a relationship to consciousness such that ‘its reality cannot be rationally doubted’ it entails that the reality of consciousness cannot be rationally doubted – just as the postulation of a ‘dormatvie power’ in sleeping pills entails that they put you to sleep – but
one might suspect that no explanation has been given here. However, the crucial point is
that Revelation isn’t a philosopher’s invention which is cooked up to try to explain Super-
Justification; if it were then the proposed explanation would indeed be rather unsatisfying.
The point is that Revelation captures how our relationship with consciousness seems to be
upon reflection. It does seem that my pain is directly present to me in such a way that its
nature is known and its reality certain.

Having said that, my argument is not: it seems to be the case that \( x \), therefore \( x \). What I am
presenting is an argument for the best explanation of Super-Justification (which is a bedrock
premise). But the fact that Revelation reflects how things seem to be, rather than being a
philosophical concoction put together for the purpose of explaining Super-Justification
removes the worry of ad hocery. By supposing that things are as they seem to be we get a
good explanation of Super-Justification; therefore, things are as they seem to be. Compare:
There seems to be an external world, and supposing that things are as they seem to be in
this respect explains the structure and regularity of our experience; therefore there is an
external world. Unless and until deniers of Revelation can give us an alternative explanation
of Super-Justification, we have strong reason to accept Revelation, as the best explanation
of Super-Justification.

Naïve realists will deny that Revelation captures how things seem to be.\(^8\) According to this
view, the properties we are immediately aware of in experience are properties of the
external world, for example, colours on the surfaces of objects for example. Experiential
properties, considered as introspectable intrinsic features of experience, are a philosopher’s
invention. As the name of the view implies, proponents of naïve realism argue that it view
captures how things seem to be; it is the naïve view of the common woman.

Of course it is in some sense true that naïve realism reflects how things seem to be upon
first reflection. But I would also say that it doesn’t take much philosophical reflection to
dissuade oneself of that viewpoint. When one goes through the process of Cartesian doubt
– doubting the external world, and then realising that one cannot doubt the reality of one’s

\(^{8}\) See, for example, Martine 1997.
own experience – one immediately realises that the properties one is aware of in experience are possibly separable, or at the very least conceivably separable, from the properties of the objects of experience. Naïve realists may claim that the process of Cartesian doubt leads us into confusion, but then the onus is on them to make a case for this. Although naïve realism is how things seem to be before one does any philosophy, it is plausible that Revelation captures how things seem to be after a little basic philosophical reflection.

Of course this brief discussion of naïve realism does not constitute anything like an adequate response to it, and I appreciate that the view constitutes a strong challenge not only to this argument, but to the very foundations of the view I defend in this book. But one cannot respond to all challenges in a single book, and I will have to defer to the criticisms others have explored in much greater detail. My aim in these two paragraphs has been simply to clarify my view as to the sense in which Revelation captures how things seem to be.

Although it is for me an epistemic building block, it seems unlikely that Revelation is a brute truth about the world. Plausibly Revelation is grounded in some more fundamental relationship which conscious subjects bear to their conscious states. Independently of the issues we have been considering in this chapter, many philosophers have defended the idea that instantiating a conscious state involves bearing a certain kind of non-mediated, non-conceptual, pre-reflective awareness relation to that state. Following philosophical custom we can call that relationship ‘acquaintance’. A creature cannot experience pain, for believers in acquaintance, without being acquainted with their pain, and in that basic sense aware of it.

Given a commitment to acquaintance, it is natural to explain Revelation in terms of acquaintance. Hamsters are acquainted with their conscious states, but lack the cognitive sophistication to enable them to attend to those states. For creatures able to attend to their conscious states, the non-conceptual awareness of acquaintance is transformed into a

---

9 For example, Robinson 1985 and 1994
10 Russell 1910 argued that we are acquainted in this sense with our sense data.
conceptual understanding of the nature of those states: Acquaintance + Attention = Revelation.\textsuperscript{11}

This is of course only a cursory account of the relationship between acquaintance and revelation; others have attempted to spell out the details in more depth.\textsuperscript{12} I mention acquaintance simply to pre-empt a concern that I am taking Revelation to be a rather strange kind of brute truth. My positive case for Revelation is not dependent on a commitment to acquaintance. Rather, it is dependent on the need to explain Super-Justification.

Can the physicalist account for Super-Justification?

Is there any way for a physicalist to account for Super-Justification? Some phenomenal concept strategists have defended a constitutional account of direct phenomenal concepts, according to which direct phenomenal concepts refer to phenomenal qualities in virtue of being partly constituted by those qualities. When I am in agony, and I deploy a phenomenal concept to think about my agony, that concept, and the thought it is involved in, are literally constituted by the feeling of agony itself.\textsuperscript{13}

We can understand this view by analogy with quotation marks. When we put quotation marks around a word, we create something that refers to a word by containing that word as a part. For example, the term ‘dog’ refers to a certain three letter word of the English language, and it does so by containing that word within itself. We can think of phenomenal concept as ‘quoting’ the phenomenal qualities they refer to, on the constitutional account. This model is not – at least not obviously – inconsistent with physicalism: the phenomenal

\textsuperscript{11} The word ‘acquaintance’ is used in lots of different ways. I am here using it specifically to mean a relationship between subjects and their conscious states which underlies and explains Revelation. For more on this notion of acquaintance see Chalmers 2003.

\textsuperscript{12} See Chalmers 2003 for a detailed account of acquaintance and its relationship with phenomenal concepts.

\textsuperscript{13} Papineau 2002, Block 2006, Balog 2012.
quality ‘quoted’ may be a physical property, even though it is not revealed as such to the concept user.

Proponents of the constitutional account claim that it is an advantage of their view that it can account for the infallibility of certain phenomenal judgements. Given that a direct phenomenal concept of pain involves within its constitution the feeling of pain it refers to. Any subject judging that she is in pain by employing a direct phenomenal concept of pain must therefore be in pain: she must be feeling pain in order for that pain to constitute the direct phenomenal concept. Such a judgement, therefore, cannot fail to be true.¹⁴

Even if this view can secure infallibility, it is not obvious what implications this has for justification. Even if as a matter of fact my belief cannot fail to be true, it doesn’t follow that I am entitled to believe it with something close to certainty. Super-Justification, as the name suggests, is a thesis about justification, not infallibility. But even if the constitutional view can ensure the super-justification of some introspective judgments, it can be applied only to a very narrow class of such judgments; too narrow as it turns out.

The constitutional account of infallibility/Super-Justification only applies to introspective judgments of the following form: I am currently having this experience, where a direct phenomenal concept is used to pick out this experience. There is no way of using the constitutional model to make sense of the super-justification of our knowledge of resemblances between experiences, such as the case discussed above of the super-justified judgement that my experience of the red splodge resembles my experience of the orange splodge.

There is a quite different approach to the phenomenal concept strategy which might be thought better able to handle these cases. I have in mind are semi-descriptivist views, according to which phenomenal concepts are a kind of hybrid concept formed of descriptive and radically opaque elements. Such views have been defended by Janet Levin and Robert Schroer. The advantage of taking phenomenal concepts to involve descriptive elements, is that these descriptive elements can potentially account for the substantive information

¹⁴ This account of infallibility is defended by Papineau (2002) and Balog (2012).
conveyed by phenomenal concepts concerning their referents. As Schroer puts it, the hope is that the involvement of these descriptive elements can account for the ‘beef’ of phenomenal concepts.\textsuperscript{15}

Levin’s view mingles elements of analytic functionalism and type-B physicalism. The reference of a phenomenal concept is determined in two distinct ways. On the one hand, a phenomenal concept picks out a conscious state by characterising it in causal-functional terms and/or in terms of the similarities and differences between phenomenal qualities. On the other hand, a phenomenal concept picks out a state by being causally connected to that state in the right kind of way. This makes reference a little precarious, as both of these reference-fixing routes must be successful for the phenomenal concept to refer. If the phenomenal concept fails to be causally connected in the right way to anything, or if the thing the phenomenal concept is causally connected to doesn’t have the relevant causal-functional/similarity properties, then reference fails.

Schroer thinks that phenomenal concepts are descriptive concepts which pick out conscious states in terms of their mode of composition from more basic phenomenal elements. The concept ‘phenomenal orange 7’, for example, characterises its referent as being composed of phenomenal red hue and phenomenal yellow hue, as well as some specific degree of phenomenal saturation and phenomenal lightness. But the sub-concepts referring to the phenomenal elements are radically opaque: reference to phenomenal elements is determined by causal facts outside of what is a priori accessible. For Schroer, the phenomenal elements turn out to refer to physical properties, and hence the conscious state out of which the phenomenal elements are composed turns out to be a composite physical state. Zombies are nonetheless conceivable, as there is no a priori information in the sub-concepts which would allow us to know that they refer to physical properties.

Both of these views are able to account for the super-justification of the following judgment: phenomenal red resembles phenomenal orange. On Schroer’s view it is

\textsuperscript{15} Levin 2002, Schroer 2010. Diaz-Leon (2014) also argues that phenomenal concepts are translucent, but does not offer an account of the kind of knowledge I’m focusing on here, i.e. resemblances between phenomenal states.
stipulated by the description that determines the reference of the concept <phenomenal orange> that the state denoted involves phenomenal red hue, and it is stipulated by the description that determines reference of the concept <phenomenal red> that the state denoted involves phenomenal red hue. In virtue of this it is a priori that phenomenal red and phenomenal orange share a common element and thus resemble, in something like the way it is evident to the understanding that ‘French women’ – so described – resemble ‘African women’ – so described – in both being women. On Levin’s view, it is built directly into the descriptive content of the concept <phenomenal red> that its referent resembles phenomenal orange, rendering it incoherent that anything correctly called ‘phenomenal red’ fails to resemble anything correctly called ‘phenomenal orange.’ In either case, <phenomenal red resembles phenomenal orange> becomes a kind of analytic truth, and hence known with comparable certainty to the truths of logic and mathematics.

The limitations of these semi-descriptivist views become apparent when we consider phenomenal resemblance not as a matter of propositions abstracted from any specific experience, but as holding between conscious states actually instantiated by the judge at the time of judgment. In the art gallery example, my super-justified judgment is not merely that of the fact that anything that is a red experience resembles anything that is an orange experience, but that this – a conscious state I demonstrate in my experience – resembles this – another conscious state I demonstrate in my experience.

Semi-descriptivism views are unable to make sense of the super-justification of these judgments. Suppose though introspection I am aware of a conscious state which presents itself as falling under the concept <phenomenal orange>. Assuming that that quality is phenomenal orange, it is a priori that it resembles phenomenal red. But how do I know that that quality I am aware of is phenomenal orange? Neither Schroer nor Levin fill in the details, but presumably there must be some reliable sub-personal mechanism that gathers information about the quality, and ensures that I deploy the correct phenomenal concept, i.e. <phenomenal orange> rather than <phenomenal green>, <pain>, or <anxiety>.

This puts introspective beliefs about which phenomenal qualities I am currently instantiate in the same epistemological boat as ordinary perceptual beliefs: reliant on the proper functioning of an information-gathering mechanism. My belief that there is a table in front
of me is reliant for its truth on the proper functioning of my senses. Similarly, for Schroer and Levin, my judgement that I am experiencing phenomenal orange (as opposed to phenomenal green, pain, or anxiety) must be dependent on the proper functioning of some information-gathering mechanism involved in introspection. Levin and Schroer cannot make sense of my judgment "I am currently experiencing phenomenal red and phenomenal orange" being super-justified, and hence – given that on their view my rational knowledge of their resemblance flows from their being correctly characterised – they cannot account for the fact that my judgment "the two states I am currently demonstrating resemble" is super-justified.¹⁶

One option the physicalist might want to explore is that phenomenal properties are a kind of response dependent property, such that all it is for me to have experience e is for me to judge that I am having experience e.¹⁷ This option would secure infallibility – I can’t be wrong that I’m experiencing e if all it is for me to experience is for me to judge that I’m experiencing e – but at the cost of a much less realist view about experience. Physicalists tend to want to identify conscious states with fully real physical properties of the brain, properties which can be identified with neurophysiological predicates, rather than mere shadows of our judgments.

Moreover, if this approach is to rule out the epistemic possibility (for the subject) of her phenomenal judgments being false, then the response dependent nature of conscious states must be apparent to the subject instantiating them. If I denote my conscious states with opaque concepts, then even if as a matter of fact my conscious states are response-dependent properties it will not be apparent to me that my conscious states are response dependent properties, and hence it will be epistemically possible for me (even though in fact impossible) that my phenomenal judgments are false. Hence Super-Justification can be

¹⁶ I am not claiming that the judgment "I am currently experiencing phenomenal red and phenomenal orange" is super-justified. The point is that for Levin and Schroer to account for the fact that the judgment "the two states I am currently demonstrating resemble" is super-justified, they would need to make sense of the judgment "I am currently experiencing phenomenal red and phenomenal orange" being super-justified.

¹⁷ Katalin Balog suggested this response to me in conversation, and it resembles the view of Shoemaker 1981, 1982.
accounted for in terms of response dependent phenomenal properties only if phenomenal concepts are transparent; and this is all that is required to secure the crucial premise of the Transparency Conceivability Argument.

**Full and partial revelation**

One might question whether revelation of the *complete essence* of phenomenal red and phenomenal orange is required to account for Super-Justification. Following Galen Strawson we can distinguish full from partial revelation:

*Full Revelation* – In having a direct phenomenal concept of token conscious state C, C is directly presented to the concept user, in such a way that (i) the complete nature of the type to which C belongs is apparent to the concept user, (ii) the concept user knows with certainty (or something close to it) that C exists.

*Partial Revelation* – In having a direct phenomenal concept of conscious state C, some aspect A of C is directly presented to the concept user, in such a way that (i) the complete nature of the type to which A belongs is apparent to the concept user, (ii) the concept user knows with certainty (or something close to it) that A exists.\(^\text{18}\)

Just as it is important to distinguish Revelation from Transparency, so it is important to distinguish Partial Revelation from mere Phenomenal Translucency; in both cases the former entails the latter but not vice versa. The views of Levin and Schroer outlined above involve Phenomenal Translucency: phenomenal concepts reveal part of the essential nature of the states they denote. But they do not entail Partial Revelation, because for Levin and Schroer translucency is secured in virtue of the semi-descriptive nature of phenomenal concepts, rather than in virtue of phenomenal concepts involving a *direct presentation* with the features outlined above.

Once clearly defined, it is apparent that Partial Revelation would not account for Super-Justification. When I form a direct phenomenal concept of conscious state C, what I am

\(^{18}\) Strawson 2006: 252. This is not exactly how Strawson defines these terms.
super-justified about is \textit{that I instantiate C}; whereas Partial Revelation entails only Super-Justification of the proposition \textit{I instantiate A}, where A is some aspect of C. Perhaps instead we might try to explain Super-Justification in terms of the following thesis:

\textit{Partial Revelation*} – In having a direct phenomenal concept of conscious state C, C is directly presented to the concept user, in such a way that (i) there is some aspect A of C, such that the complete nature of the type to which A belongs is apparent to the concept user, (ii) the concept user knows with certainty (or something close to it) that C exists.\footnote{Strawson 2006: 252. This is not exactly how Strawson defines these terms.}

By definition Partial Revelation* would entail Super-Justification. But Partial Revelation* is a cooked up notion that bears no resemblance to the revelatory relationship we to seem (upon reflection) to stand in to our conscious states. The relationship postulated in Full Revelation has two epistemic implications – the certainty that the property is instantiated and the knowledge of its nature – which are unified around the notion of direct presentation. It is because my pain is directly presented to me (it is ‘right there’ for me) that I am certain of its existence, and it is because my pain is directly presented to me (it is ‘right there’ for me) that its nature is apparent to me. If it is being directly presented with a (token of) a property which makes its nature apparent to me, and (a token of) pain is directly presented to me, then the complete nature of pain ought to be thereby made apparent to me. It makes no intuitive sense for a direct presentation of \textit{the entire token of pain} to reveal merely the nature of \textit{some aspect} of that token. And because it bears no relationship to how things seem to be, the postulation of Partial Revelation* does not constitute a good explanation of Super-Justification; it is simply the ad hoc proposal that there is some relation R such that R accounts for Super-Justification.\footnote{One might propose that we have a fully revelatory relationship of \textit{phenomenal resemblance relations}, but not their relata. Such a proposal would fail to account for the super-justification of our judgements that conscious states themselves are instantiated. But the physicalist might suppose that this element of Super-Justification could be secured for in some other way, perhaps via the constitution view. A further problem is that phenomenal resemblance relations are internal relations, grounded in the nature of the relata. It is apparent to us (with super-justification) that phenomenal red is such that of necessity it resembles phenomenal orange. Revelatory access to the phenomenal resemblance relation alone could not account for...}
Moreover, a physicalist is unlikely to want to commit to Partial Revelation – as opposed to mere Phenomenal Translucency – as it’s hard to see it could be accounted for in physical terms. Physicalist theories of reference tend to involve causal connections between concept and referent, or perhaps descriptive information concerning the referent. It’s hard to see how a revelatory relationship between a concept user and a property, a relationship in which the essence of the property is directly revealed to the understanding of the concept user, could be entirely accounted for in physical terms.\textsuperscript{21} Katalin Balog, who offers an extensive articulation and defence of the type-B constitutional view, at times sounds as though she is committing to Revelation. She uses the word ‘acquaintance’, and says that introspecting our conscious states afford us a ‘direct, unmediated, substantial insight into their nature.’\textsuperscript{22} However, she also says that ‘this kind direct insight...does not reveal anything about the metaphysical nature of phenomenality.’\textsuperscript{23} So whatever Balog means by ‘substantial’, it does not involve knowing the essential nature of experience.

\textbf{The Conceivability Argument and the Transparency Argument}

I take myself at this point to have established Full Revelation, on the grounds that a commitment to Full Revelation is the best explanation of Super-Justification (my bedrock

\begin{flushleft}
this. A final proposal: phenomenal orange is a complex property composed of aspects A1 and A2; the phenomenal concept <phenomenal orange> refers by constitution to the full complex property, which accounts for our certainty that the full property is instantiated, but also involves revelatory access to A1, which accounts for our knowledge that phenomenal orange resembles phenomenal red. In order to avoid the Specialised Transparency Argument discussed at the end of the suggestion, it would have to be argued that A1 is a purely causal-structural property. Even if this could be done, the view entails that <phenomenal orange> involves two distinct modes of presentation referring to conceivably separable properties, which in itself seems quite implausible.\textsuperscript{21} As I have discovered in conversation, it is this kind of consideration which has persuaded Joe Levine to reject physicalism, rather than modal argument (rephrase to be about full transparency). Point about the description of Revelation being transparent.  
\end{flushleft}

\textsuperscript{21} Balog 2012: 33.  
\textsuperscript{22} Balog 2012: 23.  
\textsuperscript{23} Balog 2012: 24.
commitment, and hence to have secured the truth of premise 3 of the Transparency Conceivability Argument:

**Transparency Conceivability Argument**

*Premise 1* – ‘P and nothing has Q’ is conceivably true.

*Premise 2* – *(TCP)* If a transparent sentence is conceivably true, then it’s possibly true.

*Premise 3* – ‘P and nothing has Q’ is transparent.

*Conclusion* – ‘P and nothing has Q’ is possibly true, and so pure physicalism is false.

But what about premise 2, the crucial principle allowing us to move from conceivability to possibility? In the last chapter we encountered three principles connecting conceivability to possibility: SCP, 2D-CP, and TCP. I argued by a process of elimination that TCP is the one to go for. There are plausible counterexamples to SCP: the Kripkean a posteriori necessities. And 2D-CP is a kind of TCP+, where the aspects which go beyond TCP have not been sufficiently defended. But why think that there must be any such connection between conceivability and possibility? On the face of it, what is conceivable is an epistemic matter and what is possible is a metaphysical matter. Why should the former have implications for the latter?

It is important to make vivid what is involved in denying even the weak link between conceivability and possibility asserted by TCP. A counterexample to TCP would be a strong impossibility, a thought T which cannot possibly be true, despite the fact that (A) a complete understanding of the essence of T’s truthmaker (i.e. the fact which would have to obtain for T to be true), in conjunction with (B) idealised rational capacities, would not allow one to work out that T cannot possibly be true. Suppose that T is the thought *that there is a million-sided object*. Imagine an omnipotent and perfectly rational being trying to bring about a million-sided object. She has a complete understanding of what it would be for there to be such an object, she spends an infinite amount of time examining the notion from all sides,
and can find no contradiction or incoherence in it. Hence she confidently tries to create such an object and finds...she is unable. Such an object is just brutly impossible: impossible in a way that can never be rendered intelligible, even to God. Intuitively, something seems to have gone wrong when coherence and possibility and so radically divorced from each other.

There are also good theoretical reasons for accepting TCP, as it fits well with what to my mind is the most plausible proposal philosophers have thus far dreamed up for accounting for the metaphysics of possibility and necessity: explain truths concerning possibility/necessity in terms of truths concerning essences.\(^{24}\) If the modal status of proposition P is determined by the essences of the entities P refers to, then it is natural to think that complete knowledge of those essences will allow one to work out (given sufficient power of reason) P’s modal status. This is pretty much equivalent to TCP. It is beyond the scope of this book to launch a complete defence of this approach to modality, but the fact that TCP fits well with a plausible theory of modal truth gives us reason to take it seriously.

Thus, I think a strong case can be made for TCP. It is crucial to note, however, that once we have committed to Direct Phenomenal Transparency it is not clear that we need to justify a move from conceivability to possibility in order to reject pure physicalism, or type-A physicalism for that matter. For according to either of these view my conscious states are causal-structural states. If phenomenal concepts reveal the complete nature of the conscious states they refer to, and conscious states are pure physical states, then phenomenal concepts should reveal their referents to be pure physical states. But this would entail that there is no epistemic gap between the pure physical and the experiential, contrary to what we established in chapter 3.

Thus, we can wield the following argument against pure/type-A physicalism:

The Transparency Argument against Pure/type-A Physicalism

Premise 1 – If Direct Phenomenal Transparency and either pure physicalism or type-A physicalism are true, then phenomenal concepts reveal their referents to be pure physical states.

Premise 2 – If phenomenal concepts reveal their referents to be pure physical states, then there is no epistemic gap between the pure physical and the experiential.

Premise 3 – There is an epistemic gap between the pure physical and the experiential.

Conclusion 1 – Therefore either Direct Phenomenal Transparency is false, or pure physicalism and type-A physicalism are false.

Premise 3 (Revelation) – Direct Phenomenal Transparency is true.

Conclusion 2 – Pure physicalism and type-A physicalism are false.25

I think that this kind of argument is what is really at the heart of the intuition that physicalism cannot account for consciousness. The deep conviction that pain is not a physical property is not grounded in a chain of reasoning starting from the conceivability of c-fibres firing without pain, and then moving to the real possibility of c-fibres firing without pain. It is rather grounded in the conviction that we know what pain is through feeling pain; and hence that if pain were c-fibres firing we’d know about it. But we don’t, so it isn’t. I directly grasp the qualitative nature of pain when I attend to it; and it is thereby apparent to me that what I am attending to is more than mere causal structure.

The Dual Carving objection

Dual Carving is the thesis that there are multiple ways of grasping the essence of some properties or kinds. More precisely, for at least one property/kind K, there are at least two

25 Hard Partial Revelation would also yield a specialised version of the Transparency Argument by analogy with the specialised version of the Transparency Conceivability Argument. However, at this point I am assuming Full rather than Partial Revelation.
conceptually distinct transparent concepts of K. That is to say, there are transparent concepts C1 and C2 of K, such that one could not move a priori from knowing that K is instantiated thought of under C1 to knowing that K is instantiated thought of under C2, or vice versa.

I have implicitly assumed at number of points in my discussion and defence/partial defence of the knowledge and conceivability arguments that Dual Carving is false, as I have pointed out in footnotes. The falsity of Dual Carving is also an implicit assumption in the Transparency Argument. The key move in the argument is that if direct phenomenal concepts reveal the essence of their referents, and their referents are essentially pure physical states, then direct phenomenal concepts would reveal their referents to be pure physical states. However, if Dual Carving is possible, it could be that thinking of pain under a phenomenal concept, and thinking about pain under a pure physical concept, are just two equally good ways of grasping the essence of pain.

Since I raised the issue of Dual Carving in a 2011 paper, there have been two attempts to give instances of it.26 In my 2011 paper I suggested that <bachelor> was an example of a transparent concept, and went on to say that ‘it is difficult to make sense of the thought that the notion of a bachelor…could be understood in two conceptually distinct ways.’27 This was supposed to be prima facie evidence for the implausibility of Dual Carving. However, Esa Diaz-Leon has pointed out there are in fact a number of conceptually distinct ways of understanding the property of bachelorhood:

- (A): For x to be a bachelor is for x to be an unmarried man
- (B): For x to be a bachelor is for x to be an unmarried male Homo Sapiens.
- (C): For x to be a bachelor is for x to be an unmarried Homo Sapiens with XY chromosomes.

26 In my 2011 paper I referred to the negation of Dual Carving with the clearly unbiased title ‘The Thesis of Dubious Intelligibility’ 😊
27 Goff 2011: 198.
• (D): For x to be a bachelor is for x to be an unmarried male member of such and such clade (insert here a correct description of the clade that corresponds to the species Homo Sapiens, according to the cladistic account of species).

What I think we learn from this is that I was just plain wrong that <bachelor> is a transparent concept, given that <male> is not a transparent concept. Indeed, if we suppose, as Diaz-Leon does, that the essence of biological species is partly historical, then the essence of a species will not be entirely graspable, and consequently it will not be possible to form a transparent concept of an unmarried male homo sapiens.

This is in line with my suggestion in the final section of chapter 4 that particular individuals are ‘brute one-offs’: the particular existence of a particular individual is not the kind of thing which can be grasped in thought. Consider twin tigers, the counterparts of tigers on another actual planet which is qualitatively indiscernible from our own throughout its history. On a historical account of biological species, twin tigers are not tigers, as they have different ancestors from tigers. But there is no graspable nature which differentiates twin tigers from tigers, as there is no graspable nature which distinguishes twin tiger ancestors from tiger ancestors. The ancestors of the twin tigers are just brutally numerically distinct from the ancestors of tigers.

Hence, I am now inclined to think not only that <bachelor> is not a transparent concept, but that there could not be a transparent concept of the property of being a bachelor, on account of the fact that there could not be a transparent concept of the species homo sapiens. If this is correct, then Diaz-Leon’s examples do not provide us with an instance of Dual Carving.

To be fair, Diaz-Leon claims only that A-D give us examples of multiple ‘transparent/translucent’ ways of referring to a single property, which is aimed at my claim in the 2011 paper that the physicalist cannot even hold that phenomenal concepts are translucent:28

28 Diaz-Leon 2013: 10
If the phenomenal concept of pain is translucent, then it reveals an aspect of that property. But, crucially, an aspect of a wholly physical state is itself a physical state. Therefore, if the phenomenal concept of pain were translucent, it would reveal that how pain feels involves a physical state. But this is precisely what the a posteriori physicalist denies. A posteriori physicalism is inconsistent with the claim that phenomenal concepts are translucent.²⁹

I agree that A-D are plausibly examples of translucent concepts, and I now accept that the above argument fails. The accounts of phenomenal concepts by Schroer and Levin which we explored earlier in this chapter each entail that phenomenal concepts reveal purely structural features of conscious states, structural features which may well be a priori entailed by the physical facts. This kind of translucency – what I earlier called ‘soft translucency’ – is not threat to physicalism. I think such accounts are inadequate, as I tried to show earlier, but I concede that dismissing the translucency option for the physicalist is not as easy as I had previously thought.

The second attempt to provide an instance of Dual Carving is due to John Henry Taylor.³⁰ Taylor suggests that a certain fairly popular view on the metaphysics of properties entails Dual Carving, namely the powerful qualities view, associated with C.B. Martin and John Heil.³¹ According to Martin and Heil, causal powers are neither bare dispositions nor grounded in qualitative nature; rather causal powers are identical with the qualitative properties of objects. An example often given in support of this thesis is sphericity. We may consider sphericity as a kind of quality: the quality an object has when all its surface points equidistant from its centre. But we may also consider sphericity in terms of the dispositions it endows to an object, such as the disposition to roll, or to imprint a spherical shape when pressed into plasticine.

Taylor’s appeal to the powerful qualities view is multi-faceted. Not only does he take it to provide an instance of Dual Carving, but it also suggests a novel way of defending type-B functionalism. We can identify phenomenal states with functional states, but rather than

²⁹ Goff 2011: 197.
³⁰ Taylor 2013.
embracing the standard rather implausible type-B physicalist line that phenomenal concepts tell us nothing about the nature of those states (or heaven forbid embracing analytic functionalism), we can simply say that functional concepts refer to mental states under a dispositional mode of presentation whilst phenomenal concepts refer to *those very same states* under a qualitative mode of presentation.

The problem with Taylor’s suggestion is that the very examples which the proponents of the powerful qualities view raise in support of their view also make it plausible that powerful qualities are not an instance of Dual Carving. Indeed it is this very fact which gives the view whatever plausibility it has. Martin tries to persuade us that qualities are identical with dispositions by persuading us that there are *intelligible connections* between dispositional and qualitative nature. There seems to be an intelligible connection between the categorical property of sphericity and the disposition to roll (or to make a spherical impression when pressed into plasticine), which makes us question received philosophical wisdom that we have here two properties. If the connection between sphericity-understood-qualitatively and sphericity-understood-dispositionally struck us as brute, then consideration of the case would lend no support whatsoever to the powerful qualities view. Either the powerful qualities view does not entail Dual Carving, or we have no reason to believe it.

Even if there are no clear instances of Dual Carving, why should we be reluctant to suppose that it’s true? The basic problem with Dual Carving is that, like the denial of TCP, it commits us to the world being radically unintelligible. If we deny TCP, then there are modal facts which are unintelligible even to God. And if our world admits of Dual Carving, no amount of reasoning can reveal to us that the numerous way of carving up the world ‘hang together’, i.e. are just different ways of understanding the same reality. An omnipotent and infinitely rational being could bring into being a wholly pure physical reality, and be surprised that that same reality can also be described in experiential terms. Indeed, Dual Carving implies the falsity of TCP, as there are unintelligible necessary connections between the twocarvings. Thus, even though the Transparency Argument does not involve a link from conceivability to possibility, the resistance to strong impossibilities which justifies such a link goes hand in hand with resistance to Dual Carving.
Dual Carving physicalism, as we might call it, is much more radical than type-B physicalism as it is typically understood. The natural way to understand the type-B position is that the physical truths fail to entail the phenomenal truths only because phenomenal concepts are opaque (or softly translucent). The neurophysiological or functional essence of the properties denoted by phenomenal concepts can in principle be intelligibly derived from the fundamental physical facts. In this sense there are no unintelligible connections in reality itself; the epistemic gap is wholly the result of our mental concepts failing to reveal the reality they denote. The type-B physicalist need not deny that the world would be in any respect unintelligible to a being for whom all truths were transparent.

I take the deep intuitive strangeness of Dual Carving (upon careful reflection) to be a strong reason to reject it, just as the deep intuitive strangeness of strong impossibilities is a strong reason to accept TCP. However, it is important to note in this context that my demand for intelligibility in rejecting Dual Carving is much weaker than the ‘Canberra plan’ demand of Frank Jackson and David Chalmers that all truths be a priori deducible from the fundamental truths. I am quite happy to allow that all sorts of truths may be opaque, and hence not deducible from the fundamental truths. It is only the small and very special category of transparent truths that I claim are deducible from the fundamental truths, deducible because they reveal what is required from reality for their truth: a transparent thought reveals what it would be, metaphysically speaking, for the fact upon which its truth is dependent to obtain. If phenomenal truths are not in this category then I don’t accept that we have reason to expect them to be a priori entailed by the fundamental facts.

As opposed to the across the board intelligible connections which the Canberra plan is committed to, my bedrock thesis here is a kind of minimal rationalism, a conviction that reality is minimally intelligible in the following sense:

*Minimal Rationalism (1st formulation)* – For any non-fundamental truth T, a *transparent rendering* of T is a priori entailed by a *transparent rendering* of the fundamental truths,

---

Where a ‘transparent rendering’ is defined as follows:

**A transparent rendering**: A transparent rendering of description D is a description which is indiscernible from D except that each term expressing a non-transparent concept is replaced with a term expressing a transparent concept of the same entity.

A transparent rendering leaves unchanged facts about reference and extension. It merely, as it were, opens the curtains on each concept, revealing the nature of the entity referred to.

The principle of Minimal Rationalism given above has limited application. For as discussed in (the Transparency Conceivability Principle section of) it is plausible that fundamental individuals lack graspable essences. And for any entity which lacks a graspable essence it will not be possible to think of that entity under a transparent concept; and hence in so far as a description involves terms referring to entities which lack graspable essences, it will not be possible to swap those terms for co-referring transparent concepts. If the fundamental facts involve entities without graspable essences, we will be unable to transparently render them.

However, assuming that every property and kind has a graspable essence, we can avoid this problem by formulating Minimal Rationalism in terms of *non-singular truths*, i.e. truths descriptions which don’t involve singular terms:

**Minimal Rationalism (2\textsuperscript{nd} formulation)** – For any non-fundamental non-singular truth T, a transparent rendering of T is a priori entailed by a transparent rendering of the fundamental non-singular truths.\(^{33}\)

---

\(^{33}\) Does the lack of a priori entailment between descriptions involving singular terms entail that reality is unintelligible? As far as I can see, not in any problematic sense. So long as the essence of non-fundamental individuals are defined in terms of the essences of fundamental individuals (see footnote 33 of chapter 4), we can make sense of their being grounded by analysis in facts about fundamental individuals. Slight qualification: in order to ensure that the fundamental facts *logically entail* what is required for each non-fundamental entity to be party of reality, we would need to ensure that the terms referring to fundamental entities in the metaphysical analysis of non-fundamental entities are a priori equivalent to the terms referring to those fundamental entities in the fundamental truths.
At best appeal to dual-carving looks like a particularly desperate way for the physicalist to account for consciousness. It smacks of cheating. After all, with respect to any kind of truths which are troublesome to naturalise or account for, we could pull this kind of move:

1. Suppose we are naturalists worried that the normative force of genuine reasons cannot be captured in a naturalistic vocabulary. No worries, just identify reasons with natural properties, but hold that there are two conceptually distinct ways of grasping the essence of those properties: one normative and one natural. We might even indulge in real fire and brimstone moral properties, such as sin and evil, but identify these properties with common or garden natural properties with which they bear no intelligible relationship.\(^{34}\)

2. Suppose I believe in a world of quiddities, the nature of which can be completely captured in non-causal terms, but I’m worried that without real powers the continuing regularity of the world is just a massive cosmic coincidence.\(^{35}\) No worries, just identify the quiddities with pure causal powers, even though there is no intelligible connection between conceiving of a property as a quiddity and conceiving of it as a pure causal power.

3. Suppose you really want to believe in the spirit of the woodland, who cares for the trees and wishes them well. No worries; simply hold that the chemical nature of the bark can be described in these terms.\(^{36}\)

Dual Carving seems to involve giving up any kind of constraint in metaphysics. Each of the above views smacks of cheating: smuggling in ontology without paying a fair price for it. Tax avoiders stick to the exact letter of the law whilst blatantly flouting its spirit. Dual carvers are the ontological equivalent: sticking to the exact letter of Occam’s razor but disrespecting its spirit.

\(^{34}\) I am not ruling out the view that ethical properties are identical with natural properties. My point is just that the naturalist needs to do some work to make such identities intelligible, as is done for example in Frank Jackson’s 1998 defence of ethical naturalism.

\(^{35}\) This sort of worry, discussed in chapter 9, is defended in Strawson 1986/91.

\(^{36}\) As far as I know no one in the literature has defended this specific identity ;-)
Perhaps the proponent of Dual Carving might counter this worry by imposing some kind of theoretical cost on a commitment to multiple fundamental carvings; a cost to ideology rather than ontology. But in this case, if the ‘monistic’ view that fundamental reality admits of both physical and mental carvings ends up being just as theoretically costly as standard forms of dualism, it is far from clear that it is appropriately classed as a form of physicalism.

At any rate, if the physicalist wants to try to make sense of Dual Carving, the onus is on her to make sense of this strange idea, and to argue that physicalism remains attractive with such a commitment. Until that is done, I maintain the falsity of Dual Carving, from which follows the truth of premise 1 of the Transparency Argument.

This completes my case against pure physicalism.