Revelation, intuition, and essence: an investigation into anti-physicalist thinking.

Robinson, Will

Awarding institution:
King's College London

Revelation, intuition, and essence: an investigation into anti-physicalist thinking

by William P. Robinson.

(Submitted for the MPhilStud in Philosophy at King's College London)

Word count: **30000 (inc. footnotes)**.

Table of Contents

Abstract

The present thesis has two aims. The first, which is taken up in Chapter I and will compose the majority of this thesis, is to demonstrate an important application of Revelation to the anti-physicalist project, one that is subtly distinct from the applications already discussed in the existing literature, which hitherto has not been explicitly appreciated. This demonstration is twofold: first, it is shown that several of the canonical anti-physicalist arguments are ultimately appeals to intuition, second, that Revelation entails that the intuition to which these arguments appeal - the intuition that conscious experience cannot be physical - should not be as ubiquitous as it is. In effect, this serves to demonstrate that those anti-physicalist arguments, despite being appeals to intuition, can hit their intended targets once the truth of Revelation is assumed. The second aim of this thesis, taken up in Chapter II, is to provide a more thorough understanding of the central (to Revelation) notion of 'essence.' Here, I explore two options available to the proponent of Revelation, the modal account and the real definitional account, finding flaws in both. In each case, I will argue that, in light of these flaws and their specific implications for a formulation of Revelation which adopts either of these accounts of essence, proponents of Revelation ought to look elsewhere for an appropriate account of essence. It is important to note that the *truth* of Revelation, although something that will need to be further defended in future anti-physicalist literature, lies outside the scope of the present thesis.

Over the past 15 years, the thesis of *Revelation* has emerged as the current choke-point between physicalists and their detractors. Before then, it seemed that the anti-physicalist discourse had reached a dead-end, or, at least, an impasse. In response to anti-physicalist arguments, the popular physicalist position has been to accept that there is an insurmountable gap between the physical and conscious experience, but to observe that this gap is only at the *conceptual, or epistemic,* but not *metaphysical* level. *Of course* - they will say - it does not make sense that the taste of chocolate just is the firing of certain neurons, but this need not entail that that is not the case. Responding to this, David Chalmers, in his conceivability argument, attempts to demonstrate that there cannot be a conceptual/metaphysical gap in the case of consciousness, and that if it seems that consciousness is not physical, then indeed that is the case. In this argument, however, Chalmers relies on the controversial premise that conceivability always entails some possibility, a premise which the physicalist need not accept. This is the point at which Revelation enters the debate. Revelation is the thesis that we are in a special epistemic situation with regards to our conscious experience to the extent that we are able to know our conscious experience in a peculiarly robust way. This thesis has been shown to have two broad anti-physicalist applications. First, it is able to plug the aforementioned hole in Chalmers's conceivability argument, namely its unstable foundation on the controversial premise that every conceptual possibility matches up with some metaphysical possibility, with Revelation being applied to a less controversial conceivability premise, namely that conceptual possibilities regarding conscious experience do indeed match up with metaphysical possibility. The second application of Revelation which has been explored in the literature is its more direct incompatibility with physicalism; roughly, Revelation entails that if conscious experience were physical (e.g. if my taste of chocolate just is certain neurons firing), we would know, so the fact that we do not know means that conscious experience cannot be physical. Given these applications, Revelation is therefore a significant point of contention in the mind-body debate.

The present thesis has two aims. The first, which is taken up in Chapter I and will compose the majority of this thesis, is to demonstrate an important application of Revelation to the anti-physicalist project, one that is subtly distinct from the applications mentioned above, which hitherto has not been explicitly appreciated. This demonstration is twofold: first, it is shown that several of the canonical anti-physicalist arguments are ultimately appeals to intuition, second, that Revelation entails that the intuition to which these arguments appeal -

the intuition that conscious experience cannot be physical - should not be as ubiquitous as it is. In effect, this serves to demonstrate that those anti-physicalist arguments, despite being appeals to intuition, can hit their intended targets once the truth of Revelation is assumed. The second aim of this thesis, taken up in Chapter II, is to provide a more thorough understanding of the central (to Revelation) notion of 'essence.' Here, I explore two options available to the proponent of Revelation, the modal account and the real definitional account, finding flaws in both. In each case, I will argue that, in light of these flaws and their specific implications for a formulation of Revelation which adopts either of these accounts of essence, proponents of Revelation ought to look elsewhere for an appropriate account of essence. It is important to note that the *truth* of Revelation, although something that will need to be further defended in future anti-physicalist literature, lies outside the scope of the present thesis.

Chapter I - Revelation and the intuition of distinctness.

As we will see in this chapter, many anti-physicalist arguments appeal, either explicitly or implicitly, to the familiar anti-physicalist intuition that *this*, my pain experience, cannot be *that*, the firing of my C-fibres. In response to these arguments, the popular physicalist position has been to accept that there is this intuition, but explain it away while maintaining that pain *really is* the firing of C-fibers. The primary aim of this chapter is to demonstrate that the truth of Revelation - which will be assumed as part of this thesis, for the sake of argument - blocks this physicalist response. That is, I aim to show that, according to the truth of Revelation, *there should not be this widespread anti-physicalist intuition in the first place*. The secondary aim of this chapter will then to be demonstrate how this implication of Revelation can be used to strengthen the various anti-physicalist arguments to be discussed in this chapter, all of which appeal to that intuition, and all of which, without Revelation, it will be seen, fail to hit their intended targets.

I begin (in §1) by defining physicalism, introducing the basic anti-physicalist intuition, and expositing Saul Kripke's discussion of mind-brain identities and the lessons to be drawn (and not drawn) from it regarding both physicalism and the anti-physicalist intuition. Next, I explore the anti-physicalist challenge, in its weaker form due to Joseph Levine and its stronger form due to Galen Strawson, that the non-intelligibility of mind-brain identities is a problem for physicalism, and argue with David Papineau that this challenge is really just an expression of the anti-physicalist intuition, and so lacks its intended force against the physicalist (§2). Finally, I introduce Revelation as the thesis that the essences of our experiences are *a priori* knowable to us, and argue that this thesis is able to plug the holes in the Kripke, Levine, and Strawson's arguments, given that all of these arguments appeal to the anti-physicalist intuition and that, as I will argue, Revelation entails that this intuition should not exist (§3). This discussion will highlight the importance of Revelation in contemporary and future anti-physicalist discourse, which will in turn motivate the more thorough examination of exactly how we are to think of 'essence' in formulating Revelation, which constitutes chapter II of the thesis.

§1. Physicalism, Kripke, and the intuition of distinctness.

Let *physicalism* be the view that every real concrete phenomenon, including conscious experience, is physical. There are a number of ways of understanding what is meant by 'physical' here. Naturally, 'physical' might just mean that which is posited by physics, although this understanding runs into Hempel's dilemma (1980): either we are talking about current physics, in which case we are most likely wrong about what 'the physical' refers to, given that it is highly unlikely that current physics is complete and not subject to future revision, or we are talking about an ideal, completed physics, in which case 'physical' becomes a vague term with unknown referents. To avoid this dilemma, it is common for physicalists to adopt the *via negativa* approach on which 'physical' simply means 'not fundamentally mental' (e.g. Spurrett & Papineau 1999).[1] For the sake of simplicity, I will largely focus on *mind-brain identity physicalism*, the variety of physicalism on which mental states are identified with (physical) neural states. On mind-brain identity physicalism, pain, for example, is said to be identical to the firing of C-fibres, or, to phrase this another way, pain *is* the firing of C-fibres.[2] By identifying mental states with physical neural states, the identity physicalist *reduces* the former to the latter; thinking in these terms, we can then phrase the above example claim as the claim that pain is *reducible to* C-fibres' firing. Given this narrowing of focus, I will hereafter refer to mind-brain identity physicalism simply as 'physicalism', and note explicitly when I mention other forms of physicalism; moreover, I will refer to the kind of identity statements made by this sort of physicalist simply as 'mind-brain identity statements.'

Although the 20th century arguably saw a renaissance for physicalism, it also saw the beginning of the latest wave of anti-physicalist sentiment which continues into the contemporary dialectic (although physicalism still enjoys majority support).[3] This sentiment, although now wearing slightly more sophisticated clothing, is not new. For about as long as human beings have been thinking about the nature of mind, there has been the intuition that the mind cannot possibly be reducible to the same kind of stuff that makes up rocks and

[1] We will see what 'not fundamentally mental' comes to in §2 below.

[2] That pain is C-fibres firing is an outdated and overly simplistic thesis, but this nevertheless remains the placeholder example ubiquitous in the literature, and I will follow this tradition accordingly (if only for the sake of simplicity). All objections levelled at this placeholder claim apply equally to all physicalist identity statements regarding the mind and brain.

[3] According to the PhilPapers 2020 survey, 56.5% of philosophers endorse or at least lean towards physicalism.

chairs; the prevalence of *dualism*, the view that mind and brain are distinct in a way that physicalism denies, throughout this long intellectual history is indicative of this. During the 20th century, the scope of this intuition narrowed to a particular aspect of mentality that is by far the most elusive: conscious experience. This more narrow intuition, then, goes something like this: how can *this*, the taste of chocolate, possibly be identified with *that*, the firings of certain neurons in my olfactory and gustatory cortices?[4] Following David Papineau (2002), I shall hereafter refer to this intuition as the 'intuition of distinctness.' A striking characteristic of this intuition is that it is *persistent*: even after we attend a lecture on the neuroscience of taste, in which we are informed, in as much detail as current neurophysiology can possibly afford, about the processes that lead from chocolate entering our mouth all the way up to our consciously experiencing the taste of chocolate, upon thereafter taking a bite of chocolate to verify what we have learnt, we would *still* be struck with the intuition that those processes cannot possibly *be* the conscious experience we are now undergoing. In fact, and I will return to this point a little further below, even committed physicalists must admit that there is something *funny* about mind-brain identities to the extent that *even they*, the committed physicalists, are struck with at least a semblance of the intuition of distinctness. This intuition, then, is both persistent and pervasive, and quite likely the reason that dualism and other forms of anti-physicalism will always exist amongst philosophers, scientists, and laypeople alike. Moreover, as we will see in what follows, the intuition of distinctness is at the heart of the anti-physicalist dialectic of the last half-century, with various attempts not only to vindicate it (its true vindication entails the falsity of physicalism), but also to weaponize it against the physicalist in argument. This relatively new academic interest in the intuition of distinctness and what it means for how we account for conscious experience arguably has its roots in the work of Kripke, specifically his insights regarding mind-brain identities found at the end of *Naming and Necessity* (1980).

Kripke's discussion of mind-brain identities appears in the context of his wider discussion about identity. In that discussion, Kripke establishes that whether an identity statement is necessary or contingent does not, as previously thought, correlate to whether the identity that it expresses is *a priori* or *a posteriori*, respectively. For example, 'The current Prime Minister of the UK is Rishi Sunak' expresses an *a posteriori* identity which is only

[4] The more general and oft-cited expression of this intuition is Colin McGinn's asking, 'How can technicolour phenomenology arise from sorry grey matter?' (McGinn 1989, 349.)

contingently true - after all, it is possible that Sunak could have lost the October 2022 Conservative Party leadership election and *not* thereby be the current PM.[5,6] On the other hand, however, 'Cicero is Tully' seems to express an *a posteriori* identity which is *necessarily* true. At the time of writing, this was controversial: until this time it had been thought that there were no such things as *a posteriori* necessities. But this is an extremely counterintuitive position, which we can see when considering the case of Cicero and Tully. The terms 'Cicero' and 'Tully' co-refer, they both refer to Marcus Tullius Cicero, the Roman statesman. So the statement 'Cicero is Tully' is really just saying that Marcus Tullius Cicero is identical to himself. It is hard to see how this true statement could have turned out otherwise: it might have been the case that 'Tully' became the anglicised name of another person (perhaps Servius Tullius, a legendary Roman king), but it is surely impossible that Cicero, *that very person*, could have been anybody else other than himself.[7] In fact, taking identity to be a relation between a thing and itself in this way, it seems trivial that *all* identity statements are necessarily true, if true at all, as with 'Cicero is Tully,' and it is the *contingent* identities, like 'The current Prime Minister of the UK is Rishi Sunak,' that are the puzzling ones.

Kripke explains the difference in modal status between these two kinds of identity statements by appealing to the difference between rigid and non-rigid designators. Designators are terms which refer to things; *rigid designators* are those which necessarily refer to the thing to which they actually refer (i.e. they refer to that thing in all possible worlds), *non-rigid designators* are those which only contingently refer to the thing to which they actually refer (i.e. they refer to that thing in some, but not all, possible worlds). 'Rishi Sunak,' 'Cicero,' and 'Tully,' insofar as these terms are all proper names, are rigid designators - they pick out the same person in all possible worlds. This is why 'Cicero is Tully' expresses a necessity: 'Cicero' and 'Tully' co-refer *in all possible worlds*, it is therefore impossible that Cicero - the person to which both of those terms refer - could have been anybody but himself. 'The current PM,' on the other hand, refers to Rishi Sunak in *this* world, but there are worlds in which the term, *qua* non-rigid designator, refers to somebody else, again because Rishi Sunak might not have won the 2022 party leadership election; 'The current Prime Minister of

[5] Sunak in fact ran unopposed in that leadership election, but this again could have been different.
[6] The 'is' in this statement represents an identity relation, rather than one of predication.
[7] Here I am relying on the assumption that proper names like 'Cicero' refer *directly* - I expand on direct reference (this section) and spend more time justifying this assumption about proper names (§2) below.

the UK is Rishi Sunak,' therefore, is only contingently true. The moral that Kripke draws from this exercise is that identities are necessary if statements expressing them exclusively involve rigid designators, and, furthermore, that this is true *regardless* of whether these identities are knowable *a priori* or *a posteriori*.

For Kripke, natural kind terms such as 'heat' and theoretical scientific terms such as 'molecular motion' are also rigid designators - the thing they pick out in all possible worlds is the same thing that they pick out in the actual world. So the identity statement 'heat is molecular motion,' which truly expresses an *a posteriori* identity, turns out to be necessarily true. This is particularly striking. That heat is molecular motion was a substantial empirical discovery, and this makes it tempting to think that things could have turned out otherwise, that we could have discovered that heat was some different physicochemical phenomenon. In other words, there is a feeling that 'heat is molecular motion' is contingent, an apparent possibility that it could have turned out to be false. This felt contingency makes it hard to accept Kripke's assertion that the statement is nevertheless necessary, and that our feeling of contingency is misplaced. To demonstrate that that feeling *is* misplaced, Kripke suggests that when we are imagining what *seems* to be a possible world in which heat did not turn out to be molecular motion, what we are *actually* imagining is a possible world in which the *conscious experience*, or *sensation*, that we (in this, the actual world) associate with heat is caused by something other than molecular motion. This might, for example, be a world in which the inhabitants possess a different neural structure to us and experience (what we, in the actual world, call) heat sensations caused by streaming photons:

> 'But this is not a situation in which, say, light would have been heat, or even in which a stream of photons would have been heat, but a situation in which a stream of photons would have produced the characteristic sensations which *we* call 'sensations of heat.' (*ibid*: 131; emphasis original.)

Therefore, the contingency that we are tempted to *erroneously* attribute to the statement 'heat is molecular motion' is in fact the *genuine* contingency of a nearby statement such as 'the phenomenon which causes (what we, in the actual world, call) heat sensations is molecular motion.' In other words, there isn't really a possibility that 'heat is molecular motion' did not turn out to be true (it is, rather, a necessary truth), and those in the grips of a counterintuition about this are really just thinking of the possibility that 'the phenomenon which causes heat

sensations is molecular motion' did not turn out to be true, which is a distinct and quite genuine possibility.

This discussion, about the necessity of identities involving things to which we refer via rigid designators, comes into contact with the mind-body dialectic as Kripke moves on to discuss mind-brain identities, such as the identity expressed by 'pain is C-fibres firing.' Like 'heat' and 'molecular motion,' 'pain' and 'C-fibres' are taken to be rigid designators - they have their reference necessarily, across all possible worlds. The *a posteriori* identity expressed by 'pain is C-fibres firing,' therefore, obtains necessarily if it obtains at all. Once again, however, there is a strong sense that this statement is merely contingent; for a physicalist, this is the apparent possibility that the statement, while actually true, could have turned out to be false, for anybody else, this is the apparent possibility that the statement could turn out to be false. For now, let us refer to the felt contingency of *this* identity statement - 'pain is C-fibers firing' - as an instance of the *intuition of possible distinctness*, the intuition that conscious experience and physical neural states *could be (/could have been) distinct*, and let us say, again, for now, that this is the weaker form of, and distinct from, the fully-fledged intuition of (actual) distinctness, *viz.* the intuition that conscious experience and physical neural states *are actually* distinct.[8] If an explanation were available for *this* feeling of contingency which was analogous to the above explanation for the misplaced feeling of contingency regarding 'heat is molecular motion,' it would look something like this. Those who are in the grips of the intuition of possible distinctness when thinking about the truth of 'pain is C-fibres firing' across possible worlds, while it *seems* to them that they are imagining a possible world in which pain is not the firing of C-fibres, what they are *actually* imagining is a possible world in which the conscious experience, or sensation, that we (in this, the actual world) associate with pain is caused by something other than C-fibres firing. According to this explanation, the contingency that we are tempted to *erroneously* attribute to the statement 'pain is C-fibres firing' is in fact the *genuine* contingency of the nearby statement 'the phenomenon which causes (what we, in the actual world, call) pain sensations is C-fibres firing.' In other words, on this explanation, there isn't really a possibility that 'pain is C-fibres firing' does (/did) not turn out to be true (it is, rather, a necessary truth), and those in the grip of the intuition of possible distinctness are really just thinking of the possibility that 'the

[8] As we will see a little further below, the intuition of possible distinctness turns out to just be the intuition of (actual) distinctness, given Kripke's discussion of mind-brain identities.

phenomenon which causes pain sensations is C-fibres firing' does (/did) not turn out to be true, which is a distinct and quite genuine possibility.

However, this explanation does not work because the situation involving 'pain is C-fibres firing' and the previous situation involving 'heat is molecular motion' are not quite analogous. From a metaphysical perspective, the disanalogy lies in the fact that there is no difference between pain and the *sensation* of pain, unlike heat and the sensation of heat. There is, as Kripke puts it, no mental intermediary between the phenomenon and the observer: pain *just is* the sensation or feeling of pain (*ibid*: 151). It does not make sense, as it does with heat, to say that there is the phenomenon - pain - and then there is the *sensation* which that phenomenon causes.[9] Because of this, the imagined possible world in which pain is not the firing of C-fibres *just is* the imagined possible world in which the *sensation* of pain is not the firing of C-fibres, so it cannot be the case that someone in the grips of the intuition of possible distinctness is simply mistaking the one world for the other.

From a related, epistemic perspective, the disanalogy between the felt contingency of 'pain is C-fibres firing' and that of 'heat is molecular motion' lies in the difference between how we conceptualise pain and heat, respectively. On the one hand, our (pre-theoretical) concept of heat picks out its referent - heat - via some description of heat involving (at least one of) its accidental properties, for example *being the cause of heat sensations*. This is so *despite* the fact that 'heat' is a rigid designator: 'heat' in fact refers to the same phenomenon across all possible worlds, but *we* pick out heat by, e.g., its accidental property of causing heat sensations, and so the range of possible worlds in which *we* imagine heat to exist are really the imagined possible worlds in which (e.g.) things that cause heat sensations exist. This is why we *think* we are imagining a possible world in which heat is not molecular motion, when *in fact* we are imagining a possible world in which the cause of heat sensations is not molecular motion. Our concept of pain, on the other hand, does not pick out its referent - pain - via some description of pain involving one of its accidental properties. Pain is conceptualised *phenomenally* - we pick it out by *what it is like* to undergo it, by *how it feels*, or, in more technical locution, by its *qualitative character*. Given this, and the previous

[9] Any dissenting intuitions about this can be assuaged by noting that, if there is such a thing as 'pain' that is distinct from the sensation of pain, it is the sensation of pain that is the *explanandum* of this entire exercise, and 'pain is C-fibres firing' can be adjusted accordingly to 'the sensation of pain is C-fibres firing.' I will, in what follows, continue to refer to the sensation of pain simply as 'pain.'

observation that pain *just is* the way it feels, our phenomenal concept of pain picks out its referent - pain - *directly*, by the property of being pain itself (*ibid*: 152), as opposed to picking it out *via some description*, as with 'heat,' involving one of its accidental properties. Because of this, the range of possible worlds in which *we* imagine pain to exist *really are* imagined possible worlds in which pain exists, and so when we *think* we are imagining a possible world in which pain is not C-fibres firing, we *really are* imagining that possible world, and not some other possible world in which some other phenomenon instantiating an accidental property of pain is not C-fibres firing.

Now, there are various implications that this comparison between the felt contingencies of 'heat is molecular motion' and 'pain is C-fibres firing,' and Kripke's discussion of mind-brain identities more generally, has for the intuition of distinctness and for physicalism. The most immediate and obvious upshot is,

(K) the intuition of possible distinctness (i.e. the apparent possible falsity of mind-brain identity statements like 'pain is C-fibres firing') cannot be explained away in the same way as the apparent possible falsity of statements like 'heat is molecular motion,' *viz*. by appeal to the descriptions by which we pick out terms like 'heat' (as above).[10]

Beyond this, many[11] take Kripke's discussion of mind-brain identities to constitute the following argument against physicalism:

[K1] Identity statements involving two rigid designators are necessarily true, if they are true at all;

[K2] 'pain is C-fibres firing' and all other such mind-brain identity statements involve two rigid designators and are conceivably false; so,

[K3] 'pain is C-fibres firing' and all other such mind-brain identity statements are (actually) false.

The attribution of this argument to Kripke is appropriate to the extent that [K1] and [K2] can, roughly, be taken directly from his discussion that is exposited above: [K1] is

[10] This is an uncontroversial interpretation of Kripke, who states his intention to demonstrate (K) quite explicitly at the outset of his discussion of mind-brain identities (*ibid*: 150).
[11] See the discussion on Levine (§2). C.f. Chalmers (1996; 2010).

straightforwardly the upshot of Kripke's discussion on the difference between contingently true identity statements and necessarily true identity statements, and, taking 'conceivably false' to mean 'apparently possibly false' or 'apparently contingent,' we can see that [K2] is roughly a statement of the intuition of possible distinctness. However, this argument, as it appears above, is invalid; that is, the truth of [K1] and [K2] does not entail the truth of [K3]. For the *modus tollens* to go through all of the way, it must be further assumed that the conceivable falsehood, or apparently possible falsehood, of 'pain is C-fibres firing' entails the (genuinely) possible falsehood of 'pain is C-fibres firing.' Only then does it become not true that 'pain is C-fibres firing' is necessary, and that conclusion that it is not true follows. The problem of attributing this argument to Kripke, then, is that Kripke does not appear to make this crucial assumption.[12] That analysis therefore does not, contrary to popular exegetical opinion, constitute the above argument against physicalism.

Nevertheless, Kripke does present his analysis of mind-brain identities as constituting *some* challenge for physicalism, namely the challenge to explain away the intuition of possible distinctness. Before looking at the explicit expressions of this challenge in Kripke's analysis, it is worth exploring the implications that that analysis has on the intuition of possible distinctness itself - the *explanandum* in Kripke's challenge to the physicalist - and in particular the implication that that intuition just is the intuition of (actual) distinctness. I began this discussion of Kripke by saying that it is trivial - by which I meant, it is simply common sense - that certain identities, such as 'Cicero is Tully,' are necessarily true if true at all, at least when identity is understood as a relation between a thing and itself: how can *this very thing* be anything other than itself? In fact, it is this common sense which informs Kripke's demarcation of the rigid designators (terms whose reference across possible worlds is the same as their reference in the actual world) from the non-rigid designators (terms whose reference across possible worlds differs from their reference in the actual world). For example, his justification for 'pain' being a rigid designator is simply that,

> 'if something is a pain it is essentially so, and it seems *absurd* to suppose that pain could have been some phenomenon other than the one it is.' (*ibid*: 149; emphasis mine.)

[12] In fact, there is good exegetical reason to think that Kripke did not endorse such a principle, tacitly or otherwise. See Papineau (2007).

Conversely, it is not 'absurd' to suppose that the current Prime Minister of the UK could have been some phenomenon (in this case, person) other than the one it is, and this justifies understanding 'the current Prime Minister of the UK' as a non-rigid designator. Furthermore, if it is commonsensical that the reference of certain terms (*viz.* rigid designators) across all possible worlds is the same as their reference in the actual world, then it is also commonsensical that identity statements exclusively involving these terms are necessary if true. Here is one phenomenon, *A*, for which it is absurd to suppose that it could have been some phenomenon other than the one that it actually is; and here is another phenomenon, *B*, for which it also absurd to suppose that it could have been some phenomenon other than the one that it actually is (i.e. '*A*' and '*B*' are rigid designators). To say that *A* and *B* are identical is to say *A* actually is the phenomenon *B*, that *B* actually is the phenomenon *A*, that there really is just one phenomenon here, *A*/*B*. Therefore, the identity claim '*A* is *B*' must (by the same common sense) be necessary if true: it is absurd to suppose that *A* could have been some phenomenon other than the one that it actually is, which is *B*, and absurd to suppose that *B* could have been some phenomenon other than the one that it actually is, which is *A*.

It is part of common sense, therefore, that identity statements involving rigid designators must be necessarily true if they are true at all. Psychologically speaking, this leaves no room for being committed to the truth of such statements while concurrently harbouring doubts as to their necessity; that is, there is no room to believe them while also being in the grips of an intuition that they might not have been true. In other words, if one thinks it absurd to suppose that *A* could have been some phenomenon other than the one that it actually is, which is *B* (and vice versa), it does not seem possible, again psychologically speaking, that they can be fully committed to the truth of '*A* is *B*' while also having the intuition that '*A* is *B*' might not have been true. This has a very interesting implication for the intuition of possible distinctness, which can again be seen by comparison to the felt contingency of 'heat is molecular motion.' Now, most people are committed to the truth of that statement. Those same people, however, are likely struck by the apparent possibility that heat might not have been molecular motion. How can this be so, if there is no room to simultaneously believe that heat *is actually* molecular motion yet doubt that heat *must have been* (i.e. necessarily be) molecular motion, given that 'heat is molecular motion' is of the same form as '*A* is *B*,' namely a statement exclusively involving rigid designators? We have already seen Kripke's answer: in being subject to the intuition that heat might not have been

molecular motion, we are not imagining a possible world in which heat, the phenomenon for which it is absurd to suppose that it could have been some phenomenon other than the one that it actually is, *viz.* molecular motion, is not molecular motion, but an imagined possible world in which (e.g.) the cause of heat sensations is not molecular motion. This is due to the fact that we pick out heat via some description of some accidental property like its causing heat sensations, and so the imagined possible worlds in which *we* imagine heat to exist are really those possible worlds in which that accidental property that we associate with heat exists. So we *are* able to fully commit ourselves to the truth that heat is molecular motion, because the contingency we feel regarding 'heat is molecular motion' is in fact the contingency of (e.g.) 'the phenomenon which causes heat sensations is molecular motion,' and we are just getting these identity statements muddled up because of the way by which we pick out heat. Again, though, if we *really were* imagining a possible world in which heat is not molecular motion, it is hard to see how we could really be committed to the truth that heat is actually molecular motion, given that this identity statement is necessary, and that this necessity is commonsensical.

The upshot of (K) is that the felt contingency of mind-brain identity statements cannot be made to be consistent with a commitment to their truth in the above way. In being in the grips of the intuition of possible distinctness, we really are imagining a world in which pain, the phenomenon for which it is absurd to suppose that it could have been some other phenomenon than the one that it actually is, is not C-fibres firing. This is again because of the way by which we pick out pain: our phenomenal concept of pain picks out pain directly, not by a description of some property incidentally associated with pain, and so the imagined possible worlds in which *we* imagine pain to exist really are imagined possible worlds in which pain exists. But, given that it is absurd to suppose that pain could have been some other phenomenon than the one that it actually is, our really imagining a possible world in which pain is not C-fibres firing precludes us from believing that pain *is actually* C-fibres firing, and instead entails the intuition that pain *is not actually* C-fibres firing, at least insofar as we do not believe (what are, to our own judgements) absurdities. In other words, on Kripke's analysis of mind-brain identities, the intuition of possible distinctness turns out to *just be* the intuition of (actual) distinctness, at least to the extent that these intuitions entail each other,

and so one can't have the one without having the other.[13] There are two salient upshots here: first, anybody who is struck by the apparent possible falsehood of 'pain is C-fibres firing' - which is most, if not all, people - is subject to the intuition of distinctness; and second, Kripke's challenge to the physicalist is to explain away *the intuition of (actual) distinctness*, as opposed to simply the intuition of possible distinctness. (I will hereafter speak simply of 'the intuition of distinctness' to include the intuition of possible distinctness.)

Why is this challenge significant? Well, for one thing, the fact that the intuition of possible distinctness *just is* the fully-fledged intuition of distinctness means that physicalists *themselves* are subject to that latter intuition, to the extent that they are also struck by the apparent possibility that 'pain is C-fibres firing' might have been false. Opposing the orthodox interpretation of Kripke on which Kripke is said to be levelling some kind of conceivability argument against the physicalist (as above), Papineau instead maintains that Kripke's argument is directed towards physicalists themselves, those who are already committed to the truth of 'pain is C-fibres firing'; the crucial question to them is, given that they are already committed to this identity, why does it still seem (to them) possibly false (Papineau 2007: 479)? This interpretation of Kripke fits with the analogy with 'heat is molecular motion.' The *explanandum* regarding 'heat and molecular motion' was not the *a posteriority* of that identity, the fact that the identity was a *discovery* before which there was an apparent possibility that heat was not molecular motion; rather, the *explanandum* was the fact that *despite* the universal consensus that heat is in fact molecular motion, there is *still* an apparent possibility that this might have not been the case. Likewise for 'pain is C-fibres firing': the interesting *explanandum*, the one which, according to Papineau, Kripke is challenging physicalists to explain, is the fact that *physicalists themselves* are subject to the intuition of possible distinctness, to the feeling that mind-brain identities, although actually obtaining, might not have obtained. Again, this challenge becomes particularly significant considering that the intuition of possible distinctness *just is* the intuition of distinctness: the challenge then becomes to explain why even physicalists find it so difficult to accept physicalism, given that they, like the rest of us, appear to be able to really imagine a possible world in which pain is not C-fibres firing, and that this imagined possibility entails, according to common sense, that pain is not actually C-fibres firing (*ibid*: 482 ff.). Note that this

[13] The entailment from the intuition of (actual) distinctness to the intuition of possible distinctness is trivial, given that the actuality of *p* entails the possibility of *p*.

challenge is not as strong as the [K1]-[K3] conceivability-to-possibility argument, the conclusion of which is that physicalism is false. The point here, rather, is simply that physicalism is so hard to commit to, intuitively, given the intuition of distinctness that is held by everybody, including physicalists themselves; while this is certainly something that physicalism ought to explain (*ibid*: 484), it does not, on its own, entail the falsity of physicalism.

Apart from this implicit and more specific challenge to the physicalist to explain *their* being subject to the intuition of distinctness, there are two distinct expressions of Kripke's more general challenge to the physicalist to explain away the intuition of distinctness found in *Naming and Necessity*, which bookend his analysis of mind-brain identities. He introduces the challenge in the context of Cartesian-style arguments against physicalism, and the options available to the physicalist in responding to it, in light of his discussion on rigid designators and the necessity of identities expressed by statements exclusively containing them (Kripke 1980: 145). In its canonical form, the Cartesian argument against physicalism is the argument that the mind must be distinct from the body because it is true that the mind *could have* existed without the body. Putting this in terms of conscious experience (e.g. pain) and physical neural states (e.g. C-fibers firing), where Q is some phenomenal (i.e. relating to conscious experience) truth like 'I am in pain' and P is some physical truth like 'my C-fibres are firing,'

[D1] $\Diamond(Q \wedge \neg P)$

[D2] $P \neq Q$.

In other words, it is possible that pain could exist without C-fibres firing, therefore pain and C-fibres firing must be nonidentical. Kripke's point about this argument is that the physicalist cannot respond by accepting the premise of this argument but denying the conclusion on the grounds of invalidity, by arguing that the truth of [D1] does not entail the truth of [D2]. As we have seen, given that 'pain' and 'C-fibers' are rigid designators, the identity statement 'pain is C-fibres firing' must be necessarily true if true at all. In other words, the above Cartesian-style is valid: it cannot be possible that pain could have existed without C-fibres firing. The physicalist must therefore deny [D1]. In doing this, she must explain why, although it *seems* as though [D1] is true - that is, although there is this intuition of possible

distinctness - this intuition is ultimately illusory (*ibid*: 148). This challenge to the physicalist frames Kripke's subsequent discussion of mind-brain identities, the upshot of which is (K) - that the physicalist cannot meet this challenge of explaining away the intuition of possible distinctness as illusory by appealing to some nearby, *genuinely* possible identity statement like 'the phenomenon which causes the sensation of pain is C-fibres firing' because, as we have seen, this statement and 'pain is C-fibres firing' are semantically identical.

The second expression of Kripke's challenge to the physicalist to explain away the intuition of distinctness constitutes a stronger and more general argument against physicalism, found right at the end of *Naming and Necessity*. Kripke writes,

> 'I suspect… that the present considerations tell heavily against the usual [i.e. identity] forms of materialism. Materialism, I think, must hold that a physical description of the world is a *complete* description of it, that any mental facts are 'ontologically dependent' on physical facts in the straightforward sense of following from them by necessity. No identity theorist seems to me to have made a convincing argument against the intuitive view that this is not the case.' (*ibid*: 155; emphasis original.)[14]

Here, Kripke, having established (K) just before, seems to be arguing that, given (K), there is reason to think that the intuition of distinctness cannot be explained away *at all*, and that this provides evidence for the falsity of - '*tells heavily against*' - physicalism. In other words, it is the argument that the seeming falsehood of physicalism - the intuition of distinctness - entails (or at least constitutes evidence for) the actual falsehood of physicalism (Papineau 2007: 486); it doesn't *seem* that *this*, the taste of chocolate, could be *that*, the firing of certain neurons in the olfactory and gustatory cortices in my brain, and therefore this (probably) is not the case. As Papineau (*ibid*) notes, as with the challenge that the ubiquity of the intuition of distinctness precludes anybody, including physicalists, to be totally committed to physicalism, this form of argument - *p* seems false, so *p* is false - is generally quite weak and easily blocked by an explanation as to why *p* seems false even though it is, in fact, true, and while Kripke has established (K) that the intuition of distinctness cannot be explained away by appeal to the (in this case absent) descriptions by which we pick out the terms involved in mind-brain identity statements, this is of course not the only way to explain away

[14] Here, Kripke is using 'materialism' to refer to 'physicalism,' as defined above. This is typical for the time of his writing, and I will continue to exclusively use the term 'physicalism.'

counterintuition about what is in fact true, which Kripke himself acknowledges (Kripke 1980: 155). Nevertheless, as we will see below, much of the anti-physicalist offensive, post-*Naming-and-Necessity*, is ultimately constituted by this '*p* seems false, so *p* is false'-style argument form for the specific case of conscious experience, where the '*p* seems false' is the intuition of distinctness. By the end of Chapter I, Revelation will have been demonstrated to be a promising way at strengthening this argument form against the physicalist, such that the usual response of explaining away the intuition is unavailable to her.

§2. The non-intelligibility of mind-brain identities.

I now turn to two related anti-physicalist arguments which both take the putative *non-intelligibility* of mind-brain identity statements to pose a challenge for physicalism, due to Levine and Strawson. In general terms, these two arguments involve the argumentative step '*p* doesn't make sense, but it should,' drawing differing conclusions as to what the non-intelligibility actually entails for physicalism.[15] While Levine and Strawson both mention the intuition of distinctness in the course of their discussions, both attempt to differentiate between this intuition and the non-intelligibility of mind-brain identity statements so as to avoid the charge of simply arguing from intuition; to this end, Levine claims that the intuition of distinctness is a *result* of the non-explanatoriness, cashed out in terms of non-intelligibility, of mind-brain identity statements, and Strawson invokes a notion of non-intelligibility that is not purely epistemic in nature. After careful analysis of the two arguments, it will be demonstrated that these attempts ultimately fail, Levine's because it is, in fact, the intuition of distinctness that *produces* the non-intelligibility of mind-brain identity statements, not the other way around as Levine says, Strawson's because it more straightforwardly begs the question against the physicalist with a faulty move from non-intelligibility conceived in epistemic terms, which again, as we will see, is just a consequence of the intuition of distinctness, to non-intelligibility conceived (somehow) in metaphysical terms. Ultimately, then, both arguments will be shown to just be arguing from intuition, *viz*. the intuition of distinctness, with no progress being made from the Kripke's '*p* seems false, so *p* is false'-style challenge detailed in the previous section, no special reason being given for why

[15] As we will see, the strength of these arguments vary between simply posing a problem for the physicalist to solve and straightforwardly entailing the falsity of physicalism, respectively.

the intuition that physicalism is false in any way entails that physicalism *is* false, and therefore leaves the dialectic in exactly the same position, with it being open to the physicalist to simply provide an explanation as to why, despite the fact that pain really is C-fibres firing, this fact is so intuitively difficult to grasp.

§2.1. Levine and the explanatory gap.

I begin with Levine's (1983) argument from non-intelligibility, which appeals to what Levine sees as the explanatory deficiencies of physicalist mind-brain identity statements like 'pain is C-fibres firing.'[16] Levine begins with a critique of what he takes to be Kripke's argument from *Naming and Necessity*, involving two claims:

> 'first, that all identity statements using rigid designators on both sides of the identity sign are, if true at all, true in all possible worlds where the terms refer; second, that psycho-physical identity statements are conceivably false, and therefore, by the first claim, actually false.' (*ibid*: 354.)

This is the [K1]-[K3] argument analysed in the previous section which, I argued in agreement with Papineau, is not an argument that Kripke explicitly makes in *Naming and Necessity*. However, in assessing this argument, which he erroneously attributes to Kripke, Levine correctly stipulates that intuition is an epistemic matter, and does not, on its own, have any bearing on what is metaphysically the case, that *p seeming* false does not entail that *p is* false (*ibid*: 356); this can fairly be applied to the argument which Kripke *does* make, as detailed at the end of the previous section, namely his argument that the intuition of distinctness, especially given (K), 'tells heavily against' physicalism. So Levine does not take the intuition of distinctness to straightforwardly support the conclusion that physicalism is false. He nevertheless claims that that intuition is indicative of - insofar as it is produced by - an *explanatory gap*, the existence of which does pose a challenge for physicalism. Before assessing both Levine's claimed relationship between the intuition of distinctness and the explanatory gap, and the alleged challenge that the explanatory gap poses for physicalism, I will exposit further what this explanatory gap is supposed to be, and how Levine relates it to (non-)intelligibility.

[16] Note that this is a discussion and critique of Levine's argument as it is originally found in his 1983 paper 'Materialism and qualia: the explanatory gap,' and not his 2001 book *Purple Haze*.

Recall that the difference between 'pain is C-fibres firing' and 'heat is molecular motion' that Kripke observed is (K), that the apparent possible falsity of the former (i.e. the intuition of distinctness) cannot be explained away in the same way as the apparent possible falsity of the latter, namely by appeal to the actual possible falsity of some nearby statement. Acknowledging this difference, Levine offers a further one: on the one hand, avers Levine, 'heat is molecular motion' is '*fully explanatory*, with nothing crucial left out' (Levine 1983: 357; emphasis original); mind-brain identities like 'pain is C-fibres firing,' on the other hand, 'do seem to leave something crucial unexplained, there is a "gap" in the explanatory import of these statements.' (*ibid*.) In what way is 'heat is molecular motion' fully explanatory, in a way that 'pain is C-fibers firing' is not? Mirroring Kripke's analysis of identity statements like 'pain is C-fibres firing' and his analogous analysis of identity statements like 'heat is molecular motion,' Levine diagnoses the alleged difference in explanatory power between the two kinds of statement as being rooted in how we pre-theoretically conceptualise the *explananda* - in this case pain, and heat, respectively.

As we have seen with Kripke's analysis, our concept of heat picks out its referent *indirectly*; rather than picking out heat directly, we pick it out via a description. Above, I said that this was a description of some accidental property of heat, and the example given was the property *being the cause of* (what we, in this, the actual world, call) *heat sensations*. That property, *being the cause of heat sensations*, is an example of a property that is part of the *causal profile* of heat, along with *being the cause of the rise and fall in thermometers*, *being the cause of the lift in hot-air balloons*, etc.. Typically, it is these causal-role properties, and not just accidental properties in general, which are described in indirectly-referring concepts like 'heat.' Phrasing things slightly differently than before, then, let us say that we pick out heat via a (at least partial) description of its causal profile, or causal role, pre-theoretically at least. The search for an explanation of heat, therefore, was a search for the occupier of this causal role. This is why 'heat is molecular motion' is explanatory, for Levine: given what we know about molecules from chemistry and physics, it is *intelligible* that molecular motion is the sought-after occupier of the causal role associated with heat. Not only does this provide an explanation of heat, but an *exhaustive* explanation of heat, given that the causal role of heat exhausts our pre-theoretical conception of it (*ibid*: 357). Our pre-theoretical conception of pain, on the other hand, is not exhausted by its causal profile. As we have seen, our conception of pain is phenomenal - i.e. we think about pain in terms of how it feels, its

qualitative character - and direct - i.e. our phenomenal concept of pain refers to pain directly, not via some description of its causal profile. It is true that pain *has* a causal profile - e.g., pain gives rise to the desire for it to cease; and it is also at least plausible (albeit controversial) that pain has its causal profile, or at least some of it, such as giving rise to the desire for it to cease, *necessarily*. But it is simply not the case that this causal profile is part of our pre-theoretical, phenomenal concept of pain. The search for what pain is, then, is not the search for some causal-role-occupier like the firings of C-fibres (*ibid*). This, Levine avers, is why mind-brain identity statements always and inevitably fail to be fully explanatory: it remains non-intelligible, despite what we know about C-fibres from neuroscience and physics, C-fibres should have the qualitative character associated with pain.

It should be clear by what has been said in the previous paragraph that Levine cashes out explanation, and, by extension, explanatory force, or import, in terms of intelligibility. On the subject of finding an appropriate account of explanation that complements his claim that mind-brain identity statements leave an explanatory gap, Levine writes,

> 'What we need is an account for what it is for a phenomenon to be made *intelligible*, along with rules which determine when the demand for further intelligibility is inappropriate.' (*ibid*: 358; emphasis original.)

On Levine's stipulative definition, then, for an identity statement *S* to fully explain identity *I* is for *S* to make *I* intelligible, for *I* to be made sense of; for *S* to leave an explanatory gap with regards to *I* is for *I* to be left non-intellgible by *S*, for *S* to not make sense of *I*. So, again, the statement 'heat is molecular motion' is fully explanatory, in Levine's terms, because it *makes sense*, given that we conceptualise heat by its causal role, that molecular motion - a plausible occupier of this role - *is* heat. On the other hand, 'pain is C-fibres firing' leaves an explanatory gap because, given that we conceptualise pain directly, by its qualitative character, it *does not make sense* as to why the firing of C-fibres *is pain*, given that it does not make sense as to why C-fibres firing should have *that*, or indeed any, qualitative character. As we will see further below, although Levine is cashing out explanatoriness in terms of intelligibility, the two notions can come apart, with a statement's non-intelligibility having nothing to do with its non-explanatoriness.

It is this non-explanatoriness, cashed out in terms of non-intelligibility, of 'pain is C-fibres firing,' and all other mind-brain identity statements, which, Levine claims, accounts for the intuition of distinctness, in the sense that the intuition of distinctness *is a result of* the non-explanatoriness of mind-brain identity statements:

> 'If there is nothing we can determine about C-fiber firing that explains why having one's C-fibers fire has the qualitative character that it does - or, to put it another way, if what it's particularly like to have one's C-fibers fire is not explained, or made intelligible, by understanding the physical or functional properties of C-fiber firings - it immediately becomes imaginable that there be C-fiber firings without the feeling of pain, and *vice versa*. We don't have the corresponding intuition in the case of heat and the motions of molecules… because whatever there is to explain about heat is explained by its being the motion of molecules. So, how could it be anything else?' (*ibid*: 359.)

Strictly-speaking, Levine is specifically using the non-explanatoriness of statements like 'pain is C-fibres firing' to explain the apparent possible falsity of those statements, which I dubbed the 'intuition of possible distinctness.' Here, it does not seem that Levine acknowledges Kripke's demonstration, as discussed in the previous section, that the intuition of possible distinctness collapses into the fully-fledged intuition of (actual) distinctness. It is therefore unclear as to whether Levine means to say that the intuition of distinctness proper is caused by the non-explanatoriness of mind-brain identity statements, or just that the non-explanatoriness of mind-brain identity statements causes the intuition of possible distinctness. Regardless of what exactly Levine meant, we will see below that the intuition of distinctness cannot be attributed to the non-explanatoriness of mind-brain identity statements as Levine claims. I will argue this as part of my response to Levine's claim that the explanatory gap, cashed out in terms of the non-intelligibility of mind-brain identity statements, poses a problem for physicalism, a claim to which I now turn.

I said above that Levine's argument against physicalism involves the step, '*p* doesn't make sense, but it should.' So far, we have seen Levine's argument that *p*, which, in this case, is any mind-brain identity statement, does not make sense: mind-brain identity statements like 'pain is C-fibres firing' leave an explanatory gap in the sense that they are not fully explanatory, that the identification of pain with the firing of C-fibres cannot be made

intelligible. But why should 'pain is C-fibres firing' make sense? Why should its non-intelligibility bother the physicalist? It should be noted here that Levine does not wish to establish '*p* doesn't make sense, but it should' as a general, universal principle. For example, the value of the gravitational constant *G* is a *primitive* or *brute* fact, meaning that it lacks explanation in a way that it might appear arbitrary to us. That is, the answer to the question, 'Why is *that* the value of *G*?', is not, for example, 'The value of *G* is derived from the value of some other fundamental physical constant *H*, which in turn is derived from…etc.', but simply, 'Because it is.' The appropriateness of this sort of answer to a demand for explanation is the mark of brute facts: brute facts *just are*, there is no further 'because' about them. In Levine's phrasing, such facts do not 'demand' explanation, to be made intelligible. Levine is comfortable with the existence of brute facts like the value of the gravitational constant *G*, noting that we ought to expect such facts to crop up in our investigation into fundamental reality. But Levine thinks that phenomenal consciousness, and in particular the physicalist identification of phenomenal states with physical neural states, cannot be a brute fact, that that identification demands explanation, to be made intelligible:

> '[T]he phenomenon of consciousness arises on the macroscopic level. That is, it is only highly organized physical systems which exhibit mentality… Now, it just seems odd that primitive facts of the sort apparently presented by [statements of mind-brain identity] should arise at this level of organization.' (*ibid*: 358.)

The charge here is that the posit of any brute identities at the kind of level at which conscious experience arises ought to be avoided to the extent that this would constitute an oddity, and that physicalism, given the explanatory gap, is at risk of making such a posit.

Taking stock then, Levine has argued for three claims: (i) that mind-brain identity statements leave an explanatory gap, (ii) that the explanatory gap is responsible for the intuition of distinctness, and (iii) that the explanatory gap, insofar as its existence entails the existence of brute identities at the macroscopic level of reality, poses a problem for physicalism, because the existence of such identities would be 'odd'; these claims constitute Levine's '*p* doesn't make sense, but it should'-style argument against the physicalist.[17] The

[17] Strictly speaking, Levine's argument only requires (i), which is the '*p* doesn't make sense' premise, and (iii), which is the 'but it should' premise. His using the explanatory gap to account for the intuition of distinctness, in (ii), is important, however: as we will see just below, the rejection of (ii) complements the particular rejection of (iii) made below.

physicalist may concede (i), that there is an explanatory gap in the way outlined by Levine, but can reject (ii) and (iii). I will deal with these claims in reverse order. With (iii), Levine states that it would be odd for brute identities - identities lacking in explanation, for which a demand for further explanation would be inappropriate - to arise at the macroscopic level, and, in rejecting this claim, the physicalist need only point to unproblematic brute identities which *do* so arise, and argue that statements expressing mind-brain identities are closer in kind to statements expressing these unproblematic brute identities than to what Levine dubs 'fully explanatory' identity statements like 'heat is molecular motion.' For example, Papineau (2002: 144) points to identity statements involving proper names as being more similar to mind-brain identity statements than to statements like 'heat is molecular motion' to the extent that both identity statements involving proper names and mind-brain identity statements involve directly-referring terms. In the early-to-mid 20th century, the prevailing theory of names, originating with Bertrand Russell, was descriptivism, which held that all names are really descriptions. On this view, Cicero just is a description like 'the Roman scholar who delivered the speech *Pro Quinctio*,' a description that only applies to Marcus Tullius Cicero, and 'Cicero' therefore refers indirectly, via that description. It was Kripke (1980) who first rejected this understanding of names, observing that Cicero might not have delivered the speech *Pro Quinctio* but would nevertheless have been named Cicero such that 'Cicero' would nevertheless refer to him. The alternative story that Kripke tells is that 'Cicero' refers to Cicero simply because Cicero was named 'Cicero,' and that name has been causally transmitted from person-to-person for two thousand years such that the thought in my head - 'Cicero' - is causally connected to that initial baptism in 106BC. According to this causal view of names, which is now widely accepted over Russell's descriptivism, the terms 'Cicero' and 'Tully' refer to Cicero not via some description of the man, but directly. Given this, we cannot demand a further explanation from the identity statement 'Cicero is Tully': we *can* ask, for example, why Cicero's name is anglicised as 'Tully,' but it is redundant to ask why *that one person* to whom 'Cicero' and 'Tully' both refer, directly, *viz.* Marcus Tullius Cicero, is identical to himself. That identity, therefore, is a brute identity, and perfectly benign.

The same is true, argues Papineau, for the identities expressed by mind-brain identity statements. Given that our phenomenal concept of pain, along with our scientific concept of

C-fibres firing,[18] refers directly, we cannot reasonably demand further explanation from the identity statement 'pain is C-fibres firing' as to *why* the firing of C-fibres, the state to which, according to the statement, we refer to as 'pain,' is identical to itself (Papineau 2002: 144). As Levine himself points out, the further explanation offered by 'heat is molecular motion' is the explanation as to why it is molecular motion which satisfies the description by which we refer to heat, *viz.* the description of its causal profile. It is not, crucially, an explanation as to why molecular motion is identical to itself - again, this is not a fact that needs explaining. By contrast, the only possible further explanation of 'pain is C-fibres firing,' given that there are no descriptions by which we refer to pain, would be the explanation as to why C-fibres firing is identical to itself, and insofar as this is not something that needs explaining, a demand for further explanation of 'pain is C-fibres firing' is inappropriate, just as a demand for further explanation of 'Cicero is Tully' is inappropriate. The identity expressed by 'pain is C-fibres firing' is, therefore, brute, and, just as identities expressed by statements involving proper names are perfectly benign, so too is the identity expressed by 'pain is C-fibres firing,' and all such mind-brain identities. Levine's claim (iii) that it is 'odd' that brute identities, those that are explanatorily lacking, should crop up at the macroscopic level, is therefore refuted by this analogy between identity statements involving proper names, which express benign brute identities, and mind-brain identity statements.

A defender of (iii) who wishes to salvage at least the spirit of Levine's argument, might be unconvinced by this analogy between the non-explanatoriness of mind-brain identity statements and that of identity statements involving proper names. With the above response, the physicalist has bitten the bullet, accepted (i) that 'pain is C-fibres firing' is indeed less explanatory than 'heat is molecular motion,' but maintains that this need not worry us given that there are plenty of identity statements that are relatively non-explanatory in this way, such as 'Cicero is Tully,' but whose relative non-explanatoriness does not bother us. However, the defender of (iii) might aver, we *are* bothered by the relative non-explanatoriness of mind-brain identity statements. There is, unlike 'Cicero is Tully,' a *strong feeling* that a demand for further explanation is appropriate, a yearning for that further explanation. Here we see a disanalogy between 'Cicero is Tully' and 'pain is C-fibres firing.' On the one hand, the identification of Cicero with Tully is *intelligible*: while it might not be something that needs to be explained, for which we expect no further explanation, *it makes*

[18] Presumably, all technical scientific concepts and terms refer directly, not by description.

sense that Cicero and Tully are the same person. On the other hand, the identification of pain with the firing of C-fibers remains *non-intelligible*: while it might not be something that *needs* to be explained, we nevertheless, intuitively at least, *expect* further explanation, because it *does not make sense* that pain and C-fibres firing are the same phenomenon. So, the defence of (iii) might go, while it might not be an issue that the identities expressed by mind-brain identity statements are brute in the sense that they lack explanation, such identities are nevertheless non-intelligible, they do not make sense, and *this* is still an issue, given that the other kinds of benignly brute identities that lack explanation, like 'Cicero is Tully,' are intelligible, and don't leave us with a yearning for further explanation.

This, however, should not worry the physicalist either. That there is a yearning for further explanation with regard to 'pain is C-fibres firing' but not to 'Cicero is Tully' does not contradict what has already been established: that, insofar as the analogy between mind-brain identity statements and identity statements involving proper names *does* hold, the demand for further explanation with regard to mind-brain identity statements *is in fact* inappropriate, there *is in fact* nothing left to explain, given that (due to the involvement of directly-referring terms, as with 'Cicero and Tully') the only possible *explanandum* in the vicinity of, say, 'pain is C-fibres firing,' is that C-fibers, which we sometimes refer to as 'pain,' is identical to itself, and this again is not really an *explanandum* at all. So pointing out the disanalogy that we nevertheless *feel* as if there should be further explanation with regard to 'pain is C-fibres firing,' but not with regard to 'Cicero is Tully,' only demonstrates that that feeling has nothing to do with the relative non-explanatoriness of 'pain is C-fibres firing,' given that we do not have that feeling with regards to 'Cicero is Tully' despite that statement being analogously non-explanatory (Papineau 2002: 147). Likewise for the non-intelligibility of mind-brain identity statements: pointing out the disanalogy that we can make sense of 'Cicero is Tully' in a way that we cannot make sense of 'pain is C-fibres firing' only demonstrates that that non-intelligibility has nothing to do with the relative non-explanatoriness of 'pain is C-fibres firing,' given that we can make sense of 'Cicero is Tully' despite that statement being analogously non-explanatory.

Here we see explanation and intelligibility come apart, with intelligibility being only related to the *feeling* of explanatoriness, as opposed to explanatoriness itself. It makes sense that Cicero and Tully could be the same person *despite* the non-explanatoriness of 'Cicero is

Tully,' so the non-intelligibility of 'pain is C-fibres firing,' the yearning for further explanation with regards to that statement, cannot come from that statement's analogous non-explanatoriness. So where does this non-intelligibility come from? The answer to this question is important, because it will determine whether or not the non-intelligibility of mind-brain identities poses a real issue for physicalism. Plausibly, the reason why we find it so difficult to make sense of mind-brain identity statements, why they are non-intelligible, and why we therefore yearn for further explanation, is that we have such a hard time accepting that these identities can be true in the first place. In other words, it is plausible that the non-intelligibility of mind-brain identity statements is due to the intuition of distinctness (*ibid*), that the intuition of distinctness precludes sense from being made of mind-brain identities. This is in stark contrast to Levine's claim (ii) that the intuition of distinctness is a *result* of the explanatory gap, where the explanatory gap is cashed out in terms of non-intelligibility. We can see that this is false, however, given the analogy between 'pain is C-fibres firing' and 'Cicero is Tully': both are analogously non-explanatory, and yet there is no analogous intuition that Cicero and Tully are distinct. The intuition of distinctness, with regards to 'pain is C-fibres firing,' therefore, cannot be a result of the explanatory gap that that statement leaves; in other words, (ii) is certainly false. So it remains open, therefore, to attribute the non-intelligibility of mind-brain identity statements to the intuition of distinctness, where we are now considering non-intelligibility only as being related to the *feeling* of non-explanatoriness, rather than to any non-explanatoriness itself.

With this in mind, we may return to the charge that the non-intelligibility of mind-brain identity statements poses an issue for the physicalist, even if their non-explanatoriness does not.[19] Specifically, in the words of Levine's original challenge, the charge is that it is 'odd' that we find brute identities arising at the macroscopic level, where 'brute identities' is now taken to mean identities that are non-intelligible, that do not make sense, as opposed to identities which lack explanation as in the original expression of this challenge. In response to that original challenge, I offered Papineau's point that there are identities, such as Cicero being identical to himself, which lack explanation but which are quite benign, to quell the worry that the identity expressed by 'pain is C-fibres firing,' which

[19] Strictly-speaking, this is not an argument Levine makes, as he is more interested in arguing that the non-explanatoriness (which he cashes out in terms of non-intelligibility) of mind-brain identity statements poses an issue for physicalism. In light of the failure of that argument, however, it is worth examining this alternative argument, especially as this argument is very close to Strawson's argument, which will be examined below.

is analogously non-explanatory, is somehow not benign. In response to this new challenge, that it is 'odd' that 'pain is C-fibres' expresses an identity that is non-intellgible, such that we yearn for further explanation, I offer the above stipulation that that non-intelligibility is simply a result of the intuition of distinctness: given the intuition of distinctness, the intuition that *this*, the feeling of pain, can't really be *that*, the firing of C-fibres, it is only to be expected that 'pain is C-fibres firing' is non-intellgible. That is, the non-intelligibility of mind-brain identity statements is not indicative of some problem with physicalism, rather it is indicative of the pervasive, intuitive resistance to physicalism. And, as I have said, there is no *prima facie* reason why this intuitive resistance - i.e. the intuition of distinctness - should in any way entail that physicalism is false; for all Levine has said, there is still no reason why *p*'s seeming false should entail *p*'s being false, even in the case of conscious experience.

Levine's '*p* doesn't make sense, but it should'-style argument therefore fails, given that Levine has failed to successfully argue why the non-explanatoriness of mind-brain identity statements, cashed out in terms of their non-intelligibility, or just their non-intelligibility considered separately from their explanatory power (as above), should pose a problem for physicalism. That is, Levine secures his '*p* doesn't make sense' premise, but fails to secure his 'but it should' premise, ultimately arguing instead from intuition, insofar as the non-intelligibility of mind-brain identity statements - *p*'s not making sense - is a result of the intuition of distinctness.

§2.2. Strawson and the incoherence of brute emergence.

I now turn to Strawson's (2006) argument from non-intelligibility, which appeals to the notion of 'brute emergence,' the charge here being that the physicalist commitment to statements like 'pain is C-fibres firing' constitutes a commitment to brute emergence, which is a commitment that ought to be avoided, given the incoherence that it entails.

Strawson's argument, unlike Levine's, is specifically targeted at the way in which 'physical' was defined at the beginning of the last section, namely as that which is not fundamentally mental. Given that the *explanandum* in this argument, like all of the arguments to be discussed in this chapter, is conscious experience, the physicalist claim that Strawson specifically takes issue with is the claim that the physical is that which is not fundamentally

experiential.[20] This claim, combined with the quintessential physicalist claim that every real concrete phenomenon, including conscious experience, is physical, entails that no real concrete phenomenon, including conscious experience, is fundamentally experiential. At minimum, this is a rejection of *panpsychism*, the view that the fundamental constituents of physical matter (whatever those turn out to be: elementary particles, wave-length functions, etc.), including and especially those constituting, say, the brains of organisms which are capable of undergoing conscious experience, themselves instantiate experientiality. Physicalists, however, typically hold that it is not just fundamental constituents of physical matter which lack experientiality, but *all* physical matter save for the infinitesimal proportion of that matter that constitutes the neural systems of organisms which are capable of undergoing conscious experience, and, crucially, even then, only when that matter is arranged in incredibly specific ways. For example, according to physicalism, C-fibres, along with all other kinds of neuron, are not themselves experiential; it is only until they enter the state of firing that they exhibit experientiality, insofar as that state *just is* the experiential state of pain. According to physicalism, then, pain, and experientiality in general, is an *emergent* property, meaning that the arrangement of physical matter which constitutes experiential states only has experientiality *when so arranged*, but lacks experientiality when not so arranged; in other words, experientiality emerges from physical matter (e.g. C-fibres) that lacks experientiality, plus the specific arrangement of that matter (e.g. the firing of C-fibres). It is this commitment to what I will refer to as '*Q*-emergence' which is the target of Strawson's argument against physicalism, and I will assume in what follows, with Strawson, that this commitment is essential for physicalism, as defined.

The charge which Strawson levels against the physicalist, that this commitment entails a commitment to incoherence, is not rooted in some general principle of Strawson's that emergence is an incoherent notion, a commitment to which must always be avoided. Indeed, emergence is a generally-accepted phenomenon posited in philosophy and physics in order to explain the instantiation of properties at certain levels of reality which do not occur at the more fundamental level of reality; in other words, it is the phenomenon by which the configuration of physical matter produces new properties which that physical matter does not

[20] Note that this more specific claim about the non-fundamental-experientiality of the physical is entailed by the more general claim about the non-fundamental-mentality of the physical, insofar as experientiality is a mental property.

otherwise instantiate. Strawson's charge, rather, is that *Q*-emergence *specifically* is an incoherent notion, a commitment to which must therefore be avoided. In a similar fashion to Levine, Strawson makes this point by comparison with an example of the typical, acceptable sort of emergence, and then by arguing that the disanalogy between that sort of emergence and *Q*-emergence is enough to demonstrate that the latter is incoherent. Strawson's chosen example of acceptable emergence is the emergence of liquidity from H_2O molecules, which are not themselves liquid, and their combination. This emergence is entailed by the statement 'water is H_2O': according to this statement, for physical matter to constitute H_2O molecules, when arranged in a certain way, *just is* for it to constitute water, so that matter, when so arranged, instantiates liquidity, despite lacking liquidity when not so arranged. To a first approximation, this emergence is acceptable because it is intelligible: given what we know about H_2O molecules from chemistry and physics, to echo Levine's point about the explanatory import of these sorts of scientific reductions, *it makes sense* that we get liquid when H_2O molecules are combined in that way. In other words, the property of liquidity in water is *discernible* from the properties held by, and the relationship between, the constituting H_2O molecules. To a second approximation, and this is the crucial part of Strawson's analogy, the emergence of liquidity is acceptable because the liquidity of water arises *in virtue of* the properties held by, and the relationship between, the constituting H_2O molecules. In other words, the liquidity of water is *totally dependent* on the constituting H_2O molecules and their combination. Emergence is, as Strawson puts it, 'an in-virtue-of relation' (*ibid*: 19): properties must emerge *in virtue of* the emerged-from phenomena, their properties and their combination. This is the more important way in which non-liquidity-to-liquidity emergence is acceptable, given that emergence is a metaphysical, not epistemic, notion; that is, the notion of emergence concerns the actual relationships between constituents and constitutions, not merely the intelligibility of those relationships.

Strawson's charge against the physicalist is that the emergence to which she is committed, namely the emergence of the experiential from the non-experiential, is not acceptable, that it is disanalogous to the emergence of liquidity from non-liquidity, because, Strawsons claims, it is *brute* (*ibid*: 18). Before exploring what exactly Strawson means by this, and how the allegedly physicalist commitment to brute emergence is problematic, some terminological clarification. Above, in the discussion of Levine, it was agreed that mind-brain identities are non-explanatory and non-intelligible, and in this way they are brute. There, and

this is the typical usage of the term, 'brute' was meant as an epistemic notion, describing our ability to explain or understand some phenomena, in this case mind-brain identities, as opposed to describing something about those phenomena themselves. Although, as we will see, Strawson intends on including this epistemic understanding of 'brute' in this notion of brute emergence that he uses against the physicalist, he also wishes to import a more important, metaphysical aspect to the term 'brute,' at least when applied to emergence. With this in mind, brute emergence is, as Strawson defines it, emergence where the requisite in-virtue-of relation (found in acceptable emergence, like that of liquidity), is not present. That is, brute emergence is where the alleged emergent property emerges *in virtue of nothing* about the emerged-from phenomenon, its emergence is not dependent on anything about the way the emerged-from phenomenon is; the emergence, in other words, is not dependent on any of the properties of the emerged-from phenomenon, nor on the way in which the emerged-from phenomenon is arranged. It is clear that brute emergence, defined this way, is something to which any metaphysical thesis should avoid commitment, because it entails that *anything* can emerge from *anything else*. If we allow that brute emergence is a real phenomenon that can and does happen, that emergent properties do not have to depend in any way on the emerged-from phenomena, then we cannot rule out what are otherwise *prima facie* impossibilities, such as the emergence of existence from non-existence, or emergence of the concrete from the purely abstract. Insofar as these are possibilities that ought to be ruled out, brute emergence must too be ruled out, and any metaphysical thesis that entails it must be seen to have gone wrong somewhere.

Let us grant, then, that brute emergence is indeed an incoherent notion, a commitment to which ought to be avoided, and that, *if* the physicalist were so committed, this would entail the falsity of her view. However, for the argument to go through, it must be demonstrated that the physicalist *is* so committed, as Strawson claims, and it is not clear that this is the case. The issue is that, in justifying his claim that *Q*-emergence is brute, Strawson talks almost exclusively in epistemic terms, in terms of intelligibility. For example, this is how Strawson introduces his argument for this claim,

> 'Does [*Q*-emergence] make sense? I think that it is *very, very hard to understand* what it is supposed to involve.' (*ibid*: 12; emphasis mine.)

Moreover, in his comparison between *Q*-emergence and the emergence of liquidity in water, Strawson focuses on the difference in intelligibility between the two cases. 'We can easily make intuitive sense,' avers Strawson, of the emergence of liquidity in water, as it is 'shiningly easy to grasp,' leaving 'no sense of puzzlement.' (*ibid*: 13.) He does, as I said above, state that the emergence of liquidity has the crucial in-virtue-of relation, but even this is done through the lens of intelligibility:

> '*We can see* that the phenomenon of liquidity arises naturally out of, is *wholly dependent on*, phenomena that do not in themselves involve liquidity at all.' (*ibid*; first emphasis mine, second original.)

So the disanalogy between *Q*-emergence and the emergence of liquidity in water more obviously lies in the difference in intelligibility. This is a point we have seen, however, with Levine, that, given what we know from physics and chemistry, it makes sense that molecular motion plays the causal role we associate with heat, that the combination of H_2O plays the causal role we associate with water, and that, on the other hand, mind-brain identities *do not make sense*, that, despite everything we might know from neuroscience about C-fibres, that pain just is their firing *does not make sense*. As we also saw with Levine, this point, that '*p* doesn't make sense,' needs the complementary premise, 'but it should,' in order to pose a challenge for the physicalist. So far, Strawson has only established that *Q*-emergence does not make intuitive sense, presumably for the same reasons as those given by Levine. Here it might be helpful to make a distinction between epistemically-brute emergence, emergence that does not make sense in this way, and metaphysically-brute emergence, emergence that lacks the in-virtue-of relation, and as a result makes possible all kinds of cases of emergence which we would otherwise like to rule out. We can say, then, that Strawson has so far only established that *Q*-emergence is *epistemically*-brute, the '*p* doesn't make sense' premise, and not that it is also *metaphysically*-brute. Furthermore, for the sake of complete clarity and rigour, *that an emergence is epistemically-brute does not on its own entail that it is metaphysically-brute*. Strawson, therefore, must provide further argument.

There are two broad ways in which Strawson argues for the claim that *Q*-emergence is metaphysically-brute, both of which make a fallacious appeal to that emergence being epistemically-brute.[21] The first is by analogy with putative cases of metaphysically-brute and

[21] I am critiquing these two reasons in the reverse order as they were given in the original paper.

therefore impossible emergence, namely the emergence of extensionality from non-extensionality, and the emergence of spatial from non-spatial. As I have been doing hitherto, I will extract quotes from Strawson's argument in order to analyse the language he uses and expose his implicit fallacious reasoning. On non-extension-to-extension emergence as compared with Q-emergence, Strawson writes,

> 'Well, I think this suggestion should be rejected as *absurd*. But the suggestion that when non-experiential phenomena stand in certain… relations they *ipso facto* instantiate or constitute experiential phenomena… *seems exactly on par*.' (*ibid*: 16; non-latin emphasis mine.)

Here, the comparison Strawson is drawing between non-extension-to-extension emergence and *Q*-emergence is clearly epistemic. The same is true with his second analogy, which he offers in anticipation of the reply that non-extension-to-extension emergence *isn't* absurd,

> 'My hope is that even if [one thinks] they can *make sense* of the emergence of the extended from the unextended, they won't think this about the more radical case of the emergence of the spatial from the non-spatial.' (*ibid*: 17; emphasis mine.)

Here too, clearly, Strawson is drawing attention to the fact that these cases of emergence are all epistemically-brute, they don't make sense. Nevertheless, he intends for these analogies to somehow demonstrate that *Q*-emergence is *metaphysically*-brute as well:

> 'That's why I offer unextended-to-extended emergence as an analogy, a destructive analogy that proposes *something impossible and thereby challenges the possibility* of the thing it is offered as an analogy for.' (*ibid*: 16; emphasis mine.)

The issue here is that Strawson has established that *Q*-emergence and non-extension-to-extension emergence are analogous *only insofar* as they are both epistemically-brute. Despite this, Strawson is attempting to argue as follows: both kinds of emergence are epistemically-brute, non-extension-to-extension emergence is also metaphysically-brute, so *Q*-emergence must also be metaphysically-brute. This reasoning is clearly invalid. As I said, Strawson has only given us reason to think that the two kinds of emergence are alike with regards to their intelligibility, and it would be question-begging to simply assume that they are also alike in terms of being metaphysically-brute. Strawson's

appeal to analogy here therefore does not succeed in establishing that *Q*-emergence is metaphysically-brute.

The second way in which Strawson attempts to establish that *Q*-emergence is metaphysically-brute is by somehow turning 'intelligibility' into a metaphysical notion, such that the non-intelligibility of *Q*-emergence - that is, the fact that *Q*-emergence is epistemically-brute - entails that *Q*-emergence is also metaphysically brute. The first instance of this kind of argument is found in an earlier work of Strawson, where he writes that, in order for *Q*-emergence to not be metaphysically-brute, the experiential phenomena must totally depend on the non-experiential phenomena,

> 'in such a way that the dependence *is as intelligible as the dependence of the liquidity of water on the interaction properties of individual molecules.* The alternative, after all, is that there should be total dependence *that is not intelligible or explicable in any possible physics, dependence that must be unintelligible and inexplicable even to God, as it were*.' (Strawson 2004: 69; emphasis mine.)

Citing this passage, Strawson (2003: 15) admits that this way of putting things is misleading, given that notions of intelligibility and explicability are epistemic notions, whereas his point is metaphysical, but continues to import 'intelligibility', now considered as intelligibility from the perspective of an omniscient being, in his definition of acceptable, non-metaphysically-brute emergence. In giving that definition, Strawson writes,

> 'If it really is true that *Y* is emergent from *X* then it must be the case that *Y* is in some sense wholly dependent on *X* and *X* alone, so that all features of *Y trace intelligibly back to X (where 'intelligible' is a metaphysical rather than an epistemic notion)*.' (*ibid*: 18; emphasis mine.)

Likewise, in defining metaphysically-brute emergence, Strawson writes that,

> 'emergence cannot be brute in the sense of there being absolutely no reason in the nature of things why the emerging thing is as it is (*so that it is unintelligible even to God*).' (*ibid*; emphasis mine.)

This 'metaphysical' notion of 'intelligibility' gets Strawson nowhere, however, because it is still tacitly epistemic in nature. For example, comparing again *Q*-emergence to the emergence

of liquidity (and to the emergence of a cricket team from 'eleven things that are not a cricket team'), Strawson writes,

> 'In God's physics, it would have to be *just as plain* how you get experiential phenomena from wholly non-experiential phenomena. But this is *what boggles the human mind*.' (*ibid*: 15; emphasis mine.)

The issue is that, while Strawson thinks that he is appealing purely to metaphysics when using phrases such as 'God's physics' or 'any possible physics,' he is quite clearly relying on *our* intellectual or imaginative capabilities, which is evidenced quite plainly when he writes, following from the previous quote,

> 'We need an analogy on a wholly different scale [to the emergence of liquidity] if *we are to get any imaginative grip* on the supposed move from the non-experiential to the experiential.' (*ibid*; emphasis mine.)

When imagining what is intelligible from the perspective of God, we are limited by what is intelligible *to us*. The truth is that we do not know what is intelligible from the perspective of an omniscient being, because we ourselves are not omniscient. So building this 'metaphysical' notion of 'intelligibility' into the definitions of acceptable emergence and metaphysically-brute emergence does one of two things; either it puts us in no position to know when an emergence is metaphysically-brute, or it tacitly appeals to our own imaginative capabilities in determining what is metaphysically-brute. The first option entails that we have no way of determining whether *Q*-emergence is metaphysically-brute. The second means that Strawson is again relying on the fact that *Q*-emergence is epistemically-brute and hoping that this alone convinces us, without argument, that this entails that it is also metaphysically-brute. Given that Strawson's entire argument is aimed at *demonstrating* that *Q*-emergence is *metaphysically*-brute, and that, throughout this argument, Strawson's language is almost exclusively *epistemic*, it is clear that he means the latter, that, with this talk of intelligibility from the perspective of God, or 'God's physics,' he means to surreptitiously take *our* intellectual capabilities and venerate them in order to draw metaphysical conclusions.

In the end, Strawson has framed this entire exercise as an attempt at *understanding* *Q*-emergence, which clearly we cannot do, and then concluding that, because of this, it must

be metaphysically-brute. This argument therefore takes the form of, '*p* doesn't make sense, but it should,' in a similar fashion to Levine's argument. Both arguments begin, in their respective '*p* doesn't make sense' premises, with some non-intelligible physicalism claim - Levine targets mind-brain identity statements, and Strawson targets *Q*-emergence (which is entailed by mind-brain identity statements and the *via negativa* definition of the physical, as explained above). However, where Levine gave an explicit (albeit uncompelling) argument for his 'but it should' premise, namely that it is 'odd' that we find non-explanatory or non-intelligible identities at the macroscopic level, Strawson is not at all explicit as to why the non-intelligibility of *Q*-emergence should bother the physicalist, instead using a question-begging argument by analogy, and ambiguous use of the notion of 'intelligibility,' in order to tacitly move from the epistemic-bruteness of *Q*-emergence to the metaphysical-bruteness of *Q*-emergence. With both of these faulty tactics, it seems that what Strawson is in fact relying upon, is the intuition of distinctness, which, as stipulated at the end of the discussion of Levine, is responsible for the non-intelligibility in the first place: the reason why we cannot make intuitive sense of mind-brain identity statements, or of *Q*-emergence, is because *we already intuitively think* that mind-brain identity statements are false, that *Q*-emergence is impossible. That is, Strawson's argument by analogy, and his attempt at building 'intelligibility' *into* the definition of metaphysical-bruteness, are both really just appeals to intuition: in the argument by analogy, Strawson's hope was that we find those allegedly analogous cases of emergence *intuitively* implausible, reject them, and then reject the case of *Q*-emergence on the basis that it is analogously intuitively implausible; in taking 'intelligibility' to be intelligibility from the perspective of God, Strawon's hope was that the intuitive implausibility of *Q*-emergence would preclude us from imagining *Q*-emergence to be intelligible even to an omniscient being. In fact, near the beginning of his discussion, Strawson even *acknowledges* that his argument will involve appeals to intuition (*ibid*: 9). Moreover, in what appears to be a recreation of Strawson's argument, Sam Coleman, who seems to endorse that argument, writes

> 'For how could we obtain items for which there is an answer to the question 'What is it like?' (Answer, for instance, 'pink') *just* by rearranging items for which the answer to this question is: 'Qualitatively? Nothing at all'. How could *all there is* to the quality of the sum be but the relationships holding between qualityless items?' (Coleman 2015: 74.)

This sort of rhetoric is the textbook mark of an argument from nothing but intuition. Elsewhere, Coleman refers to what he sees as the impossibility of *Q*-emergence as a 'glaring truth' which physicalists try to evade (Coleman 2016: 250, n.5.). It is often the case that when philosophers start invoking 'glaring truths' they are simply digging their heels into the ground, banging their fists on the table, and thrusting forward strongly held intuitions which, however glaringly true they might seem to the invoker, are unsupported by real argument. And it is safe to say that this is an example of just that. The trouble is that the intuition to which Strawson and Coleman make appeals is the intuition of distinctness, the very intuition which the physicalist will explain away. We have therefore made no progress from Kripke's '*p* seems false, so *p* is false'-style argument, with no special reason as to why, in the case of conscious experience, *p*'s seeming false really does entail that *p* is false.

§3. Instrumentalising intuition: Revelation and the transparency of phenomenal concepts.

I now turn to the thesis of Revelation which, as I will have demonstrated by the end of this section, provides that special reason as to why, in the case of conscious experience, *p*'s seeming false really does entail that *p* is false, and, furthermore, is able to plug the hole in each of the three previous anti-physicalist arguments, instrumentalising the intuition of distinctness to which those arguments, as I have shown, appeal.

Although the first explicit mention of 'revelation' with regards to conscious experience was due to Mark Johnston (1992: 223), the spirit of Revelation can be traced back further to the writings of Russell and his notion of acquaintance. Acquaintance is the proposed relation which holds between subjects and certain items whereby such items are known to the subjects to whom they are related in some direct and fundamental way, more direct and fundamental than, say, any thoughts that subjects might have about those items: rather than forming a mental state 'that is (merely) *about* something, when we are acquainted with something we are, in some sense, supposed to consciously confront that very thing itself.' (Raleigh 2019: 2; emphasis original.) To confront acquainted-with items *directly* here means to become aware of such items in a way that does not require, for example, any kind of inference (Russell 2001/1912: 25; Coleman 2019: 51). For Russell, *qua* sense-data theorist, we are never acquainted with concrete phenomena out in the mind-independent world, but only sense-data, memories, other inner states like propositional attitudes, and universals (i.e.

properties) that colour these items. A striking characteristic of the kind of knowledge gained via acquaintance with these internal mental states (and their properties) is that it is in some sense complete. On experiencing colour, for example, Russell writes, 'I know the colour perfectly and completely when I see it, and no further knowledge of it itself is even theoretically possible.'(2001/1912: 25.) I say 'in some sense complete' because, presumably, Russell does not intend to mean that knowledge by acquaintance delivers knowledge of accidental features of mental states, like the time at which it occurs; rather, we can take Russell to mean that knowledge by acquaintance delivers full and complete knowledge as of the *essential* features, or simply the *essence* of inner states.[22] It is this revelatory aspect of Russell's knowledge by acquaintance to which the spirit of Revelation can be traced; it is a kind of proto-Revelation.

Although sense-data theory, of which Russell was a proponent, declined in popularity over the 20th century, the idea that we have this kind of unbridled epistemic access to our own conscious experiences has endured and indeed flourished in the recent anti-physicalist literature. For example, Strawson seems to have in mind Russell's idea, that, when experiencing colour, we come to know the colour completely and perfectly, when he writes,

> 'whatever they are, colour words are words for properties whose essential nature as properties can be and is fully revealed in sensory (and indeed visual) experience, given only the qualitative character that sensory (visual) experience has.' (1989: 213.)

This expression of Revelation is far closer to the letter of the contemporary formulations than Russell's musing, with terms like 'qualitative character' and 'essential nature.' It is this expression which Johnston (1992: 223) christens 'Revelation' a few years later. For now, let us state Revelation, as it is understood in the contemporary literature, in its most general and basic form.

(Revelation) In introspecting an experience, either occurrent or one that is held in memory or the imagination, the essence of its phenomenal properties is *a priori* knowable to the subject.

[22] Russell does not explicitly make this qualification regarding essence. Exactly what we mean by 'essential' and 'essence' will be taken up in Chapter II.

By ‘phenomenal properties,’ I mean the properties that constitute the qualitative character of an experience, i.e. how that experience feels, for example the *paininess* of a pain experience or the *redness* of a reddish visual experience. It is taken for granted in most of the literature of Revelation that phenomenal properties are internal and intrinsic to the subject, I, for the sake of ease, I will follow in this assumption. This puts certain formulations of Revelation, for example Bill Brewer’s (2019) naïve realist account of Revelation, outside the scope of this thesis. Furthermore, as I have disclaimed at the beginning of the present thesis, I will not be arguing for the *truth* of Revelation. Rather, for the rest of this section, I will be discussing the implications of Revelation, and, in Chapter II, I will be discussing what proponents of Revelation ought to mean by ‘essence.’

It is worth noting, before moving onto more robust formulations of Revelation, and the way in which they are able to instrumentalise the intuition of distinctness, that Revelation employs ‘*a priori*’ in a peculiar way as knowledge gained through introspection, where *a priori* knowledge is usually taken to be knowledge gained through non-empirically-informed reasoning. The reason this is peculiar is because knowledge gained through introspection is synthetic knowledge, making Revelation a thesis about synthetic *a priori* knowledge, a peculiar kind of knowledge - traditionally there is *a priori* analytic knowledge (e.g. Alfred Pennyworth’s knowledge that Bruce Wayne *qua* bachelor is also an unmarried man) on the one hand and *a posteriori* synthetic knowledge (e.g. Alfred Pennyworth’s knowledge that Bruce Wayne is Batman) on the other. There is precedent to treat introspection as a form of *a priori* knowledge, however, in debates surrounding semantic externalism and the self-knowledge of thought. On semantic externalism, the meaning of my thought ‘water’ constituently involves H_2O given that that’s what my thought ‘water’ refers to. This view leads to strange consequences when taken in conjunction with the view that we can know our own thoughts introspectively: if I can know my thoughts introspectively then I can know purely by introspection that I am in a world with H_2O. At this point in the debate it is put that one cannot know *a priori* about the world, and therefore either that semantic externalism is false or the view that we can know our own thoughts introspectively, in the way just outlined, is false. Here, then, introspection is considered a source of (synthetic) *a priori* knowledge. I take this as sufficient precedent for employing this non-standard understanding of ‘*a priori*’ in Revelation.

I now turn to two of the more robust formulations of Revelation commonly cited in the contemporary literature. The first is due to Philip Goff (2017), who formulates Revelation in terms of phenomenal concepts. Recall that a phenomenal concept is a concept which picks out a conscious experience *directly*, by how it feels, by its qualitative character. Phenomenal conceptualisation is generally how we conceptualise conscious experience, especially when thinking about our own conscious experience. For example, although one might have *a* concept of pain which picks pain out by some description of a causal-role property, like *causing the desire for it to cease*, when asked to think about *their own pain*, that person will almost certainly form a phenomenal concept. This would be especially true if they were undergoing a pain experience *at that moment*. In that case, they would be said to form a *direct phenomenal concept*, meaning that the phenomenal concept that they are forming is about an occurrent conscious experience to which they are attending. With this in mind, here is Goff's formulation of Revelation:

> '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 (as a token of that type).' (*ibid*: 107; emphasis original.)

Applying this to pain: in forming a direct phenomenal concept about their occurrent pain experience, the concept user is in a position to know the complete nature, which we can take to be the essence, or essential nature, of that type of pain experience, and is certain as to the occurrence of that (token) pain experience.

Goff utilises this formulation of Revelation in order to plug what is widely considered to be a hole in David Chalmers's anti-physicalist argument from conceivability (1996; 2010). Kripke showed that we can imagine a possible world in which pain is not the firing of C-fibres, and that imagined possible world really is a world in which *pain*, not just some accidental property associated with pain, is not the firing of C-fibres. Chalmers argument is that that imagined possible world is not a *merely* imagined possible world, but a genuine, metaphysically possible world; in other words, that the conceivable falsity of 'pain is C-fibres firing' entails the possible falsity of 'pain is C-fibres firing.' Given that 'pain is C-fibres firing' must be necessarily true if true at all, this possible falsehood therefore entails that

'pain is C-fibres firing' is not true at all. Recall that this is the argument that is often erroneously attributed to Kripke in *Naming and Necessity*:

[K1] Identity statements involving two rigid designators are necessarily true, if they are true at all;

[K2] 'pain is C-fibres firing' and all other such mind-brain identity statements involve two rigid designators and are conceivably false; so,

[K3] 'pain is C-fibres firing' and all other such mind-brain identity statements are (actually) false.

I said in §1 that this argument, in this form, is invalid, and that what it is missing is some principle which says that the conceivable falsity of 'pain is C-fibres firing' entails the possible (and therefore actual) falsity of 'pain is C-fibres firing.' This is what Chalmers provides with his two-dimensional conceivability principle, that

(2D-CP) If a sentence is conceivably true, then its primary intension is true at some possible world.[23]

According to Chalmers's two-dimensional semantic framework, the primary intension of a term is the reference of that term across possible worlds when that term is conceived under its descriptive content: the primary intension of 'heat,' therefore, is the reference of 'heat' across possible worlds to phenomena that, for example, cause heat sensations. The secondary intension of a term is the reference of that term across possible worlds when that term is conceived *directly*, by what it is, in essence: the secondary intension of 'heat,' therefore, is the reference of 'heat' across possible worlds to molecular motion. This is how Chalmers accounts for the fact that 'heat is molecular motion' is conceivably false yet necessarily true: while it is not possible that heat, in its essence, i.e. molecular motion, could have been anything other than molecular motion, there exists a genuine, metaphysically possible world at which the primary intension of 'heat is molecular motion,' *viz.* 'the phenomenon that causes heat sensations is molecular motion,' is false, and this is what we're imagining when we say that there is a felt contingency to 'heat is molecular motion,' that that statement is conceivably false. Generalising from this case, Chalmers argues that there is always this link

[23] I follow Goff (2017: 88) in phrasing things this way.

between conceivability, primary intension, and possibility. The final step is another of Kripke's insights, that our phenomenal concept for pain is directly referring, with no descriptive content, and so does not have distinct primary and secondary intensions; the same is true for our concept for 'C-fibres.' Therefore, the conceivable falsity of 'pain is C-fibers firing' entails that there is a metaphysically possible world in which *pain*, not some property accidentally associated with pain, is not *C-fibres firing*, not some property accidentally associated with C-fibres firing. With this, the *modus tollens* of the above argument goes through, and [K3] is secured.

The issue is that (2D-CP) is controversial, and rests on the assumption that all conceivably false, but necessarily true, identity statements have distinct primary and secondary intensions; in other words, that such statements involve at least one term that does not refer directly, but by description. It is at this point where physicalists (e.g. Papineau 2002; Loar 1990) give examples of such cases. For example (Papineau 2002: 89ff.), we might imagine somebody, Jane, who has over the course of her life picked up the names 'Cicero' and 'Tully' separately and absentmindedly to the extent that she does not remember when, where, or from whom she picked these names up. But they nevertheless exist in her mind. Jane does not have any descriptions or further ideas attached to these names, they are just maximally simple nodes in her head that happen to refer to one person, Marcus Tullius Cicero, and refer to him *directly*, without referring to him by some associated description. Now, given this directness and maximal simplicity, Jane does not know that these mentally-stored names co-refer, and so might happen to entertain the thought 'Cicero is not Tully,' and perhaps even believe that this is the case. There is no possible world corresponding to this thought, because Cicero just is Tully, necessarily. But there is also no possible world at which the primary intension of 'Cicero is Tully' is false, given that these concepts refer directly and therefore that 'Cicero is Tully' has no primary intension that is distinct from its secondary intension, and, again, the secondary intention of 'Cicero is Tully' is not true at any possible world. This serves as a counterexample to (2D-CP).[24]

So [K1]-[K3] is once again left without a conceivability-to-possibility principle; this is the aforementioned putative hole in Chalmers's argument from conceivability. The way in

[24] Chalmers (2010: 170ff.) responds to this and a number of other counterexamples to (2D-CP), but I will leave that particular debate there in order to move on to what is, regardless of Chalmers's defence of (2D-CP), a far less controversial conceivability-to-possibility principle, due to Goff.

which Goff seeks to plug this hole is by offering a new conceivability-to-possibility principle, one that does not rest on such a strong assumption, and that is not vulnerable to the above sort of counterexample. First, let a concept be *transparent* if it reveals all of the essential properties of its referent; for example, the concept 'sphericity' is transparent because it is part of that concept that for something to be a sphere every one of its points must be equidistant from its centre, which is the essence of being a sphere (Goff 2017: 15). A sentence is transparent if all of the concepts involved are transparent. With this notion of transparency, Goff gives his own conceivability-to-possibility principle, that

(TCP) If a transparent sentence is conceivably true, then it is possibly true (*ibid*: 100).

This principle is advantageous over (2D-CP) for two reasons. First, it is far more intuitive: if you grasp everything there is about the referents of the terms in a sentence, and you can conceive of that sentence being true, then how could the truth of that sentence be nevertheless *impossible*? (2D-CP), on the other hand, is not motivated by intuition, so much as theoretical economy: it is more parsimonious, argues Chalmers, to not have two distinct spaces of logically possible worlds, the 'merely conceivable' possible worlds and the genuine, metaphysically possible worlds, but to just have one space of logically possible worlds, all of which are metaphysically possible (Chalmers 2010: 187.) Of course, this parsimony is only advantageous if there are no counterexamples to (2D-CP), which we have seen is not the case. This is the second reason why (TCP) is advantageous over (2D-CP): it does not rest on the extremely strong and contentious assumption that *all* conceivably false, necessarily true identity statements must involve at least one concept that does not refer directly, meaning that (TCP) is not vulnerable the above example of Jane, for instance. It does not matter that Jane's thought 'Cicero is not Tully' is conceivably true, yet not possibly true, because neither of the concepts involved in that sentence are transparent. In fact, given that these concepts exist in Jane's head as maximally simple nodes, they are, in Goff's terms, *radically opaque*, meaning that they reveal *no* properties about their referent. In order to plug the hole in the [K1]-[K3] argument, then, Goff appeals to (TCP) along with his formulation of Revelation, which entails that phenomenal concepts are transparent. Given that our concept of 'C-fibres' are also, presumably, transparent, this entails that the conceivable falsity of 'pain is C-fibres' entails the statement's possible, and therefore actual, falsity. The *modus tollens* goes through,

and [K3] is secured, this time with a far less contentious conceivability-to-possibility principle.

Goff also devises a novel anti-physicalist argument from Revelation (2017: 147-149), but, for the sake of brevity, I won't explore that here, although I will reference it while formulating my own argument that Revelation entails that there shouldn't be an intuition of distinctness. In order to make that argument, I must first introduce a second formulation of Revelation, a formulation which will also serve as the springboard for my discussion on essence in Chapter II, this time due to David Lewis. Lewis (1995) gives this formulation in the context of investigating whether physicalism (or 'materialism', as Lewis prefers) is compatible with the notion of *qualia*. For present purposes, I will simply define 'qualia' (singular: 'quale') as the phenomenal properties which constitute the qualitative character of an experience (see §1).[25] Lewis believes that this notion is actually a part of folk psychology, because 'when philosophers tell us very concisely indeed what they mean by 'qualia', we catch on' (*ibid*: 140), and so his investigation is into 'qualia' as a folk-psychological concept, and whether this concept is compatible with physicalism. According to Lewis, the above definition that I have given is part of that concept, as well as, what he calls, the *Identification Thesis*, which Lewis formulates as follows:

> 'when I have an experience with quale *Q*, the knowledge I thereby gain reveals the essence of *Q*: a property of *Q* such that, necessarily, *Q* has it and nothing else does.' (*ibid*: 142.)

This is straightforwardly a statement of Revelation, and I will henceforth refer to what Lewis calls the Identification Thesis as Revelation. It is this aspect of the allegedly folk-psychological concept of 'qualia,' for Lewis, which makes the concept incompatible with physicalism. This amounts to the claim that, regardless of whether it is part of the folk-psychological concept of 'qualia,'[26] Revelation contradicts physicalism, such that a commitment to both is inconsistent. Lewis gives his reasoning for this:

[25] There is also a definition on which qualia are *defined* as being non-physical; given that this is a thesis about Revelation, *qua* anti-physicalist siege engine, I will not define qualia this way either, lest I render Revelation a moot point.

[26] The question of whether Revelation is a part of folk psychology is outside the scope of this thesis. See, however, Stoljar (2008) and (Liu (2021).

> ‘If, for instance, Q is essentially the physical property of being an event of C-firing, and if I identify the qualia of my experience in the appropriate ‘demanding and literal’ sense [this is how Lewis understands revelation], I come to know that what is going on in me is an event of C-firing. Contrapositively: If I identify the quale of my experience in the appropriate sense, and yet know nothing of the firing of my neurons, then the quale of my experience cannot have been essentially the property of being an event of C-firing.’ (Lewis 1995: 142.)

The reasoning here is straightforward enough. Revelation entails that, through introspecting an experience, I ought to know everything essential about the qualia which constitute the qualitative character of that experience. Presumably, physicalism wishes to say that those qualia are essentially physical properties. Therefore, Revelation entails that I ought to know, from introspection, that the qualia of my experience are physical properties. Given that I *do not* know this, Revelation entails that those qualia are not essentially physical properties.

In the rest of this section, I aim to demonstrate another way in which Revelation can work against physicalism: namely, that Revelation, conceived under the Lewisian formulation in particular, is able to instrumentalise the intuition of distinctness against the physicalist, such that Revelation entails that *there should not be such a widespread intuition of distinctness*. This will in turn, similarly to Goff’s utilisation of Revelation to plug the hole in Chalmers’s argument from conceivability, demonstrate how Revelation is able to plug the hole in each of the anti-physicalist arguments which have been discussed in this chapter, *all* of which were shown to collapse into appeals to that intuition. While, as we saw, those arguments ultimately fail to challenge physicalism in the way in which their proponents intended, where Kripke, Levine, and Strawson *succeed* is in *making vivid* the existence of the intuition of distinctness, the *seeming* falsehood of p. I will elucidate the way in which Revelation entails that the seeming falsehood of p, in the special case of conscious experience, entails the actual falsehood of p, thus allowing the arguments of Kripke, Levine, and Strawson to all secure their intended conclusions. I begin with how the truth of Revelation is incompatible with the widespread subjection to the intuition of distinctness, before applying this finding to each of the aforementioned anti-physicalist arguments.

The argument for the entailment from Revelation to the claim that there should not be such a widespread intuition of distinctness is quite straightforward, and we need only recall

what we have already learnt from Kripke and the above formulations of Revelation. Given that it is an argument about what Revelation entails, we may begin by assuming the truth of Revelation, the thesis that the essence of an experience, that is either occurrent or held in memory or imagination, is *a priori* knowable to the subject. Next, recall from the discussion of Kripke, and then of Chalmers just above, that identity statements with descriptive content can be read either as being statements about inessential appearance properties, or essence. For example, 'heat is molecular motion,' in virtue of 'heat' picking out heat via a description of, say, the accidental causal-role property, that we associate with heat, of causing heat sensations, can be read as 'the phenomenon which causes heat sensations is molecular motion,' which, intuitively, is not a statement about what heat *is*, *in essence*. The reading of 'heat is molecular motion' that *is* a statement of essence would be reading it (trivially) as '*what heat is*, *viz.* molecular motion, is molecular motion.' This is, recall, how Kripke accounts for the felt contingency of 'heat is molecular motion.' As we also saw, 'pain is C-fibers firing' does not work this way, because it lacks descriptive content, and can only be read as '*what pain is*, *viz.* (according to this statement) C-fibres firing, is C-fibres firing,' meaning that the felt contingency of this statement - the intuition of possible distinctness, as I termed it - really is a case of us imagining a possible world in which *pain*, *that very phenomenon*, and not simply some accidental property that we associate with pain, is not C-fibres. So 'pain is C-fibres firing' is a statement about the essence of pain. Given the truth of Revelation, which we have assumed for the sake of this argument, this means that 'pain is C-fibres firing' ought to be knowable *a priori*. That is, if 'pain is C-fibres firing' is true, given Revelation, and given that we think of 'pain' directly, in terms of how it feels, which presumably means either introspecting an occurrent pain, or remembering a previous experience of pain and holding that memory in our imagination, it ought to be *a priori* knowable that the essence of pain is C-fibres firing.[27,28]

But of course, 'pain is C-fibres firing' is *not a priori* knowable. This is demonstrated by the Kripkean observation that 'pain is C-fibres firing' is conceivably false. In the earlier discussion of Kripke I cashed out conceivability in terms of imagined possible worlds: to say that 'pain is C-fibres firing' is conceivably false is to say that we can imagine a possible

[27] There is a subtle distinction here between knowing *that* the essence of pain is C-fibres firing, which is how I am phrasing things, and knowing *of* the essence of pain, which is C-fibres firing. See Liu (2019) on the importance of making this distinction in formulating Revelation against the physicalist.
[28] Goff (2017: 124) makes a similar link between Revelation and the *a priority* of physicalist claims.

world in which pain is not the firing of C-fibres. Another way to put conceivability is in terms of *a priority*: to say that 'pain is C-fibres firing' is conceivably false is to say that 'pain is C-fibres firing' is not *a priori*, that we cannot *a priori* rule out the falsity of 'pain is C-fibres firing.' (Chalmers 2010: 143.) For example, the falsity of 'a sphere is a shape whose points are all equidistant its centre' is inconceivable as it can *a priori* be ruled out; likewise, the falsity of 'heat is molecular motion,' when conceived as a statement about essence, namely 'molecular motion is molecular motion' can be *a priori* ruled out. These two statements are *a priori* knowable. On the other hand, the falsity of 'pain is C-fibres firing' cannot *a priori* be ruled out, 'pain is C-fibres firing' is not *a priori* knowable. So far, this is more or less the Lewisian demonstration that Revelation and physicalism are incompatible: Revelation says that 'pain is C-fibres firing' ought to be knowable *a priori* if true, 'pain is C-fibres firing' is *not* knowable *a priori*, therefore Revelation is incompatible with the truth of 'pain is C-fibres firing.' From here, though, a further point can be made regarding Revelation's compatibility with the existence of the intuition of distinctness.

To make this point, let us begin with the intuition of possible distinctness. This was the intuition that mind-brain identity statements like 'pain is C-fibres firing' are possibly false, that it is possible that pain might not be (/not have been) the firing of C-fibres. This intuition, as with the apparent possible falsehood of 'heat is molecular motion,' was analysed in terms of imagining, or conceiving of, possible worlds: in being subject to the intuition of possible distinctness, we *think* that we are imagining a possible world in which pain is not C-fibres firing. As Kripke demonstrated, given the way by which we pick out pain (and C-fibres), we, in being subject to the intuition of possible distinctness, *really are* imagining a possible world in which pain is not C-fibers firing. We can therefore say, in being subject to the intuition of possible distinctness, that 'pain is C-fibres firing' is conceivably false to us, that we cannot *a priori* rule out the possibility that pain might not be (/might not have been) the firing of C-fibres, and that, therefore, 'pain is C-fibres firing' is not *a priori* knowable to us. The truth of Revelation, therefore, given that (as above) it entails that 'pain is C-fibres firing' must be *a priori* knowable, also entails that we should not be subject to the intuition of possible distinctness, given that being so subject is demonstrative that 'pain is C-fibres firing' is *not a priori* knowable. This again is enough to show that Revelation is incompatible with physicalism, given that we *are* subject to the intuition of possible distinctness, but my aim here is to show, further, that Revelation entails that we should not be subject to the intuition

that Kripke, Levine, and Strawson all appeal to in their anti-physicalist arguments - *viz.* the intuition of (actual) distinctness. This final step is extremely simple: intuition of possible distinctness *collapses into* the intuition of distinctness, and to that extent they can be taken to be the *same intuition*. That is, to be subject to the intuition of possible distinctness *just is* to be subject to the intuition of (actual) distinctness. This is because, as we saw in our discussion of Kripke, it is part of common sense that identity statements like 'pain is C-fires firing' (*qua* statements involving two rigid designators) are necessarily true if true at all, and this, combined with the fact that we really are imagining a possible world in which pain is not C-fibres firing when subject to the intuition of possible distinctness, means that we cannot be fully committed to the truth of 'pain is C-fibres firing': in equivalent terms, the intuition of possible distinctness just is the intuition of (actual) distinctness. The point that Revelation entails that we should not be subject to the intuition of possible distinctness therefore extends to the intuition of (actual) distinctness as well: given the truth of Revelation, there simply should not be the widespread intuition of distinctness.[29]

The general application of this point is that it strengthens the '*p* seems false, so *p* is false' style of argument in the case of conscious experience. Given the truth of Revelation, we should not be subject to the intuition of distinctness, *p* should not seem false in this way. Conversely, given the truth of Revelation, the fact that we *are* subject to the intuition of distinctness, the fact that *p does indeed* seem false in this way, entails that *p* is false. We are now in a position to see how this point can be specifically applied to each of the anti-physicalist arguments we have discussed in this thesis, all of which appeal, either explicitly or tacitly, on the intuition of distinctness, beginning with Kripke (§1). This is the argument found right at the end of *Naming and Necessity*, which is the most straightforward expression of the argument form '*p* seems false, so *p* is false' of the three arguments under discussion, simply arguing that, given (K), that the intuition of distinctness cannot be explained away in the usual way by appealing to descriptive content in mind-brain identity statements (because there is none), the existence of the intuition of distinctness 'tells heavily against' physicalism (Kripke 1980: 155). The weakness of this argument, as we saw, is the weakness of '*p* seems false, so *p* is false'-style arguments in general: opponents can maintain

[29] Liu (2021) makes a similar link between Revelation and the intuition of distinctness, arguing that the intuition of distinctness *comes from* what she argues to be the intuitiveness of Revelation. Goff (2017: 125) echoes this point in passing.

that *p* is true while providing an explanation as to why it is nevertheless intuitive. With Revelation, though, this option is blocked: according to the truth of Revelation, as I have demonstrated just above, there should not be this widespread intuition in the first place. Kripke's argument is therefore strengthened by Revelation.

Recall that Levine and Strawson's arguments (§2), while ultimately appeals to intuition, make their respective appeals more tacitly. Levine's original charge against the physicalist was that mind-brain identity statements are not explanatory in a way that, for example, 'heat is molecular motion' is, and that this makes the identities expressed by those former sorts of identity statements 'odd,' given that they arise at the macroscopic level. Drawing on Papineau's analogy with identity statements involving proper names, which are similarly non-explanatory, yet express benign (that is, not odd) identities, it was shown that the non-explanaoriness of mind-brain identity statements should not worry the physicalist. At this point, it was suggested that it is the *non-intelligibility*, rather than the non-explanatoriness, of mind-brain identity statements which makes the identities they express 'odd' in a way that ought to bother the physicalist. Here, though, it was stipulated that this non-intelligibility is just a result of the intuition of distinctness, and so ought to be expected, given the ubiquity of that intuition. So Levine's argument, which ought to abandon the charge that the non-explanaoriness of mind-brain identity statements is a problem for physicalism, instead must rely on the appeal to the non-intelligibility of such statements, which is really just an appeal to the intuition of distinctness (given that the former is really just a result of the latter). Revelation, however, entails that we should not be subject to this intuition of distinctness. So, if it really is that intuition which blocks the intelligibility of mind-brain identity statements, then Revelation also entails that mind-brain identities *should* make sense, just as 'Cicero is Tully' makes sense, and that, therefore, the non-intelligibility of mind-brain identity statements *does*, in the end, pose a problem for physicalism. Levine's argument (or, at least, the next best alternative to it) is therefore strengthened by Revelation.

Given the ways in which Revelation strengthens both Kripke and Levine's arguments, it ought to be clear how it also strengthens Strawson's argument, which appeals to emergence. Strawson argues that the physicalist, in being committed to the emergence of the experiential from the non-experiential, is committed to (metaphysically-)brute emergence, which entails that physicalism, insofar as it is so committed, is false. As we saw, Strawson fails to

demonstrate that physicalists *are* so committed, instead simply appealing to the fact that the emergence of the experiential from the non-experiential is *non-intelligible*. In a similar diagnosis to that of Levine's argument, I concluded that Strawson is therefore appealing to the intuition of distinctness, insofar as that intuition accounts for the non-intelligibility of experiential emergence (as we saw, Strawson seems to make this assessment himself). Once again, Revelation not only implies that this intuition to which Strawson appeals should not exist (or, at the very least, should not be as ubiquitous as it is), but that the non-intelligibility that the intuition produces should not exist either. In particular, in the case of emergence, it implies that experiential emergence *should* be intelligible, just as the emergence of liquidity from H_2O molecules is intelligible, a comparison that, recall, Strawson makes. Strawson's argument, therefore, is strengthened by Revelation.

* * *

In this chapter, I have demonstrated the importance of Revelation to the anti-physicalist project, not only as a source of independent argument against the physicalist, but also as a way to strengthen the many existing anti-physicalist arguments which, without Revelation, are simply appeals to intuition which do not otherwise pose a particularly strong challenge for the physicalist. Given this importance of Revelation, future dialectic around Revelation ought to be centred around (i) whether Revelation is true, and (ii) how we are to best understand Revelation. For the rest of this thesis, I will begin to explore (ii); specifically, I will begin to explore how proponents of Revelation, especially given the importance of Revelation to the anti-physicalist project, ought to understand 'essence.'

Chapter II - The essence of conscious experience.

It is common to most formulations of Revelation - robust or not - to speak of the 'essence' or 'nature' of qualia, or experiences.[30] The aim of this chapter, then, is to explore the options available to the proponent of Revelation. Also common to these sorts of formulations of Revelation is the vagueness with which this notion (essence) is deployed - it is usually the case that philosophers making these formulations have a preferred understanding of essence, but this is rarely made explicit in or around the formulation itself. It should therefore be instructive to explore various accounts of essence in order to find the most appropriate for Revelation, especially given its importance to the anti-physicalist dialectic, which was demonstrated in Chapter I. I begin (in §4) with the Lewisian formulation of Revelation and the account of essence that it appeals to, the modal account, arguing, with Liu (2019), that this account is inappropriate for Revelation. Next, I explore the alternative account suggested by Liu, the real definitional account, arguing that it too is flawed and therefore not an account of essence which proponents of Revelation should adopt (§5). As a result of these critiques, it will ultimately be left open as to how proponents of Revelation ought to understand 'essence,' with the conclusions of this chapter being entirely negative.

§4. Lewisian essence as necessity.

Recall that, in Lewis's understanding of Revelation, Revelation is the thesis that introspecting an experience whose qualitative character is constituted by quale Q reveals the *essence of* Q. Lewis, for his part, subscribed to the once-popular modal account of essence, as evidenced in his above formulation of Revelation:

> 'the essence of *Q*: a property of *Q* such that, *necessarily, Q has it and nothing else does*.' (*ibid*: 142; emphasis mine.)

On the modal account of essence, an essential property of *a* is simply a necessary property of *a*, and the essence of *a* is the sum of all such properties which is sufficient for being *a*. Here, the entailment between a thing and its essence goes both ways: in all possible worlds, if you're F then you're also G, and, in all possible worlds, if you're G then you're also F. This is

[30] See, e.g., Strawson (1989), Johnston (1992), Lewis (1995), Goff (2017), and Liu (2019).

because of the proviso '*and nothing else does*' -- necessarily, F things have G and *nothing else does*. The entailment between a thing and its essential properties, on the other hand, only goes one way: *having my parents* is an essential property of mine, but if I had any siblings, they would also have that essential property, and therefore it isn't true that, in all possible worlds, if somebody has my parents, then that person is me. So *having my parents* is an essential property of me but not the essence of me. Take another example, regarding squares. All squares have the property of *having four equal straight sides* and the property *having equal internal angles*. Both of these properties, individually, are essential properties of squares, but neither one (again, taken individually) is the essence of squares because rhombuses (i.e. diamonds) have four equal sides but unequal internal angles , and all sorts of regular polygons have equal internal angles but aren't four-sided. The *conjunction* of these two properties, however, is sufficient for a thing being a square, and therefore the essence of squares.

Applying this understanding of essence to talk about the essence of an experience: necessarily, experiences with the qualitative character constituted by quale Q have the essence X *and nothing else does*. Note that this application of the modal account differs slightly from Lewis's, as he is quoted above, at least. In that application, the essence of Q is described as a *property of* Q,[31] a second-order property, given that 'Q' - *qua* quale - is itself a property. For Lewis, qualia are properties of experiences (*ibid*: 142), where experiences are considered as events (*ibid*: 141); Liu, in her revised Lewisian formulation, follows Lewis in defining qualia this way, *viz.* as properties of particular events of experiencing (Liu 2019: 229). This differs slightly from my own above definition of qualia, in which I do not specify exactly what qualia are properties *of*, only that they are internal to the subject and constitute the qualitative character of an experience. Nevertheless, on my understanding of qualia, as well as on Lewis and Liu's, qualia are *properties*, and so, given that 'essence' here is also understood as denoting a property, it is (at best) bad ontological housekeeping to formulate Revelation as the thesis that, in introspecting an experience with quale Q, I thereby gain revelatory knowledge with regards to the *essence of Q*. The awkwardness of applying this understanding of essence to Revelation in this way is likely due to the fact that that understanding - the modal understanding of essence, as defined above - works best when considering essences of

[31] Something like: the property of instantiating the conjunction of all of a thing's necessary properties, where the instantiation of that conjunction of properties is both necessary and sufficient for being that thing.

particular things, as opposed to the essences of properties. In the example application of the modal understanding of essence I gave above, I said that the conjunctive property of *having four equal straight sides and having four equal internal angles* was the necessary and sufficient property, and therefore, on that understanding of essence, the essence of *squares*. It might have seemed natural to say that that conjunctive property was the essence of *squareness* or *being a square*, but this would be to make the same mistake as Lewis appears to have done, namely that of invoking second-order properties. Similarly, with regards to Revelation, we cannot speak of the essences *of qualia*, which amounts to speaking of the essences of *having a certain qualitative character*. It is for this reason that, on my application of the modal understanding of essence to Revelation, it is not Q - some quale - that is said to have the essence X, but the thing that *has* Q, namely the experience itself, which are particulars - events, as Lewis and Liu think.

Liu (2019: 231) also takes issue with Lewis's apparent appeal to second-order properties in his formulation of Revelation, but takes this as secondary motivation for formulating Revelation with a different understanding of essence in mind, namely the older, Aristotelian real definitional account, as revived by Kit Fine (1994). Applying this to Revelation, Liu proposes to understand revelatory knowledge regarding Q as 'knowing some truth' about Q, namely that 'Q is X', where the 'is' here is taken to mean something like 'is defined by.' (Liu 2019: 232.) This, Liu avers, avoids talk of second-order properties as 'X' is now simply a predicate which 'captures the essence' of Q (*ibid*), as opposed to X being itself the essence which Q instantiates, as Lewis suggests.[32] I agree that formulating Revelation with the real definitional account of essence, as above, can avoid talk of second-order properties in the way that Liu suggests. However, as I have demonstrated, it is possible to formulate Revelation without abandoning the modal understanding of essence which Lewis prefers, by speaking of 'X' as being the essence (i.e. the conjunctive property of having all necessary properties that together are sufficient for being the thing in question) of experiences - *qua* particulars - that have Q, as opposed to the essence of Q - *qua* property - itself. That Lewis himself invokes second-order properties in his formulation of Revelation is not, therefore, motivation to abandon the modal account of essence and instead look at something like Fine's real definitional account.

[32] I explore the real definitional account of essence in detail, along with its flaws, and its applicability to Revelation, below in §5.

Such an abandonment is nevertheless well-motivated, given the flaws in the modal account of essence as laid-bare by Fine (1994); it is these flaws which Liu (2019: 231) takes as primary motivation for looking instead to the real definitional account of essence in formulating Revelation. The general claim that Fine makes in his attack on the modal account of essence is that that account, which identifies essential properties with necessary properties, gives incorrect sufficient conditions for what essential properties are. Recall that, on the modal account, essential properties are simply identified with necessary properties, entailing that a property F is an essential property of α if and only if it is also a necessary property of α.[33] Against this, Fine provides numerous counterexamples of necessary properties that are not essential properties, contradicting the 'if' part of the above biconditional statement that is implied by the modal account.[34] The first of these examples concerns the relationship between Socrates and the property *belonging to the set whose sole member is Socrates*. Presumably, given that, necessarily, singleton <Socrates> (i.e. the set whose sole member is Socrates) exists if Socrates exists, and also that, necessarily, he belongs to singleton <Socrates> if both he and the singleton exists; together this implies that the property *belonging to the set whose sole member is Socrates* is a necessary property of Socrates. In other words, there is no possible world in which Socrates exists and does not belong to singleton <Socrates>. On the modal account of essence, this implies that *belonging to the set whose sole member is Socrates* is an essential property of Socrates, i.e. is *part of the essence* of Socrates - a counterintuitive implication (Fine 1994: 4).

Each of the counterexamples Fine gives against the modal account of essence follow this first one in structure, namely having a counterintuitive implication that F is an essential property, or part of the essence of, α. Fine's second counterexample concerns the essences of two seemingly distinct objects, such as Socrates and the Eiffel Tower. Presumably, it is necessary that Socrates and the Eiffel Tower are distinct - that is, there is no possible world in which Socrates and the Eiffel tower are the same object. On the modal account, this counterintuitively implies that it is part of the essence of Socrates that he is distinct from the Eiffel Tower and part of the essence of the Eiffel Tower that it is distinct from Socrates -

[33] N.b. Fine's attack on the modal account focuses on the essences of particular things, not of properties; this supports my earlier suggestion that the modal account is only fit for dealing with the former.
[34] Fine does (1994: 8) accept the 'only if' part of the biconditional - that is, he agrees with proponents of the modal account that it is a necessary condition of being an essential property (e.g. of α) that it is also a necessary property (of α) (see §5 below).

counterintuitive, because although it is reasonable to say that the essence of Socrates and that of the Eiffel Tower are unconnected, this fact does not seem to be *included* in these essences (*ibid*: 5). Fine's final set of counterexamples concern the relationship between objects and necessary truths, and are in particular counterexamples to the implication of the modal account that necessary truths are identical to essential truths. The difficulty here is that necessary truths follow from *anything*, because they are true no matter what. For example, that there are infinitely many prime numbers necessarily follows from the proposition 'Socrates exists', given that it is itself a necessary truth. In other words, it is a necessary truth about Socrates that, if he exists, there are infinitely many prime numbers. On the modal account, this implies that it is also an *essential* truth about Socrates that, if he exists, there are infinitely many prime numbers - in other words, that it is part of the essence of Socrates that there are infinitely many prime numbers. This of course cannot be the case.

The last of these counterexamples against the modal account can be amended so as to demonstrate in particular why it ought not be applied to Revelation. According to Revelation, we are able, via introspecting an experience, to know the essence of that experience, or of the properties that (at least partly) constitute the qualitative character of that experience, *viz.* the qualia of that experience. Given that, as we have just seen, on the modal account of essence, all necessary truths are part of the essences of all things, Revelation as conceived under the modal account implies that, *from introspection alone*, we are able to know the essence of that experience / the qualia of that experience *and therefore* also to know not just *a priori* necessities like 'there are infinitely many prime numbers,' but also *a posteriori* necessities like 'water is H_2O' and 'gold had the atomic number of 79.'[35] This would make Revelation too strong a thesis. Just as, as we saw in §3, the semantic externalist must avoid the implication that my introspecting my thought 'water' ought not allow me to know that I am in a world with H_2O, the proponent of Revelation must avoid the implication that my introspecting my (say) visual experience as of water allows me to know that water is H_2O.

[35] See Kripke (1980). Note that the *a posteriori* necessities given as examples here are, according to the modal account of essence, facts about essence. C.f. Fine:

> 'Among the necessary truths, if our modal theorist is to be believed, are statements of essence. For a statement of essence is a statement of necessity and so it will, like any statement of necessity, be necessarily true if it is true at all. It follows that it will part [*sic*] of the essence of any object that every other object has the essential properties that it is: it will be part of the essence of the Eiffel Tower for Socrates to be essentially a person with certain parents, let us say, or part of the essence of Socrates for the Eiffel Tower to be essentially spatiotemporally continuous. O happy metaphysician! For in discovering the nature of one thing, he thereby discovers the nature of all things.' (Fine 1994: 5.)

That Revelation *has* this implication when taken with the modal account of essence demonstrates therefore that that account of essence is inappropriate for formulating Revelation.

The move to look to a different account of essence with which to formulate Revelation is therefore well-motivated, due to the flaws of the modal account taken on its own merit as well as when applied to Revelation. However, as we will see, what is seemingly the obvious alternative has deep flaws of its own, again both taken on its own merit and when specifically applied to Revelation, making it too an unsuitable account of essence for formulating Revelation.

§5. Essence and 'what a thing is.'

As I said, Liu prefers the real definitional account of essence for Revelation; Fine also turns to this account in light of the failures of the modal account demonstrated by his counterexamples. The real definitional account consists of two broad claims, both of which will be criticised in this section. Having completed his (successful) attack on the modal account, Fine is keen to clarify that he does not wish to completely sever whatever intuitive tie there is between essence and necessity that led philosophers like Lewis to the conclusion that these concepts denote the same thing. On this, Fine writes,

> 'Certainly, there is a connection between [essence and necessity]. For any essentialist attribution will give rise to a necessary truth; if certain objects are essentially related then it is necessarily true that the objects are so related (or necessarily true given that the objects exist).' (*ibid*: 8.)

Here Fine is homing in on the particular aspect of this intuitive connection between essence and necessity that is most intuitive, namely the entailment from essence to necessity. It does seem that it is built into our commonsense concept of essence that it has modal consequences - e.g. that it is part of my essence that I have my parents surely entails that *it can't have been the case that I* did not *have my parents*. So the modal account, to the extent that this understanding of the relationship between essence and necessity is correct, was fairly close to the mark in analysing the essentialist attribution 'it is part of my essence that I have my

parents' as the statement 'necessarily, for all *x*, if *x* has my parents then *x* is me'. The mistake made by proponents of the modal account was *analysing* the former *as* the latter, where in fact the latter is simply an *entailment* of the former. As we saw above, the entailment does not go the other way: it is true that, necessarily, for all *x*, if *x* has the property of belonging to the set whose sole member is Socrates, then *x* is Socrates, but this does not entail that belonging to the singleton Socrates is a part of the essence of Socrates.

This intuitive aspect of essence and necessity is at the core of the first of the aforementioned two claims made by Fine as part of his real definitional account of essence, namely the claim that *all necessary truths are true in virtue of the essences of some entities.* Fine often puts this in terms of necessary truths being 'sourced' in essences. The most simple case is where a necessary truth about *x* is sourced in the essence of *x*, e.g. that necessarily I have my parents is straightforwardly sourced in the essential property of mine (i.e. the part of my essence) that I have my parents. Then there are cases where a necessary truth about *x* is sourced in the essence of some other entity *y*, e.g. that necessarily Socrates belongs to the set whose sole member is Socrates is sourced, not in the essence of Socrates as the modal account implies, but in the essence of the singleton Socrates. Finally, there are cases where a necessary truth about *x* is not sourced in the essence of any particular entity, but are nevertheless sourced in the shared essence of some class of entity:

> 'For each class of objects, be they concepts or individuals or entities of some other kind, will give rise to its own domain of necessary truths, the truths which flow from the nature of the objects in question. The metaphysically necessary truths can then be identified with the propositions which are true in virtue of the nature of all objects whatever.' (*ibid*: 9.)

On this understanding of necessities, logically necessary truths 'flow' from - i.e. are sourced in - the essence of all logical objects (whatever those are), conceptually necessary truths flow from the essence of all concepts, and so on. What Fine calls 'metaphysical necessity' or 'necessary *simpliciter*' is the most general kind of necessity, and therefore metaphysically necessary truths flow from the essence of all entities - concrete, logical, conceptual, etc.. Fine's account of essence therefore seeks to source *all necessities* in (some) essence, which constitutes a commitment to *essentialism about modality*, the view that modality can be

explained in terms of essence (as opposed to the modal account which sought to explain essence and modality in terms of each other by equating them).

The second claim of Fine's real definitional account - the claim from which the account gets its name - is that *essentialist attributions* (i.e. saying *F* is part of the essence of *x*) *just are definitions*. Fine first introduces the connection that he holds to exist between essence and definition through an analogy between necessity and analyticity. Analytic truths are true in virtue of certain terms that are involved in expressing them: e.g. 'all bachelors are unmarried men' is analytically true given the meaning of the terms 'bachelor'. By giving a definition for these terms, therefore, one is able to demonstrate the analyticity of these sorts of truths. Analogously, according to Fine's essentialism about modality at least, necessary truths are true in virtue of the essences of certain entities. By giving the essence of those entities, therefore, one is able to demonstrate the necessity of these sorts of truths. Giving a definition therefore works similarly to attributing essence, in that the former functions in demonstrating analytic truth as the latter functions in demonstrating necessary truth (*ibid*: 10). It is here that Fine finds his foothold to argue from the analogy of definition and essence attributions to the identity of the two. Not only, avers Fine, does giving the definition of the terms 'bachelor' function *similarly* to giving the essence of some entity, but it is in fact a *case* of giving the essence of some entity, namely some linguistic entity. Fine initially muses that the linguistic entity which is the subject of the definition *qua* essence-attribution is the word 'bachelor' itself. This would require a conception of words on which the word 'bachelor' is partly constituted by its meaning, which would amount to the meaning of 'bachelor' being a part of its essence. In defining 'bachelor', we are therefore giving (part of) the essence of the word insofar as we are giving its meaning. However, this conception of words is controversial, with the more common - albeit less natural, according to Fine - view being that words have their meaning contingently, as a result of convention or specification, and so the meaning of 'bachelor' is not part of the essence of the word, merely an accidental feature. On this view, then, giving the definition of 'bachelor' is not an essence-attribution. Given this, Fine instead suggests that the linguistic entity which is the subject of the definition *qua* essence-attribution of 'bachelor' is the *meaning* of the word 'bachelor.' The reasoning is as follows. In giving a definition of the word 'bachelor,' we are specifying its meaning. Not all candidate specifications will be appropriate; the one that *is* will be the one which specifies what the meaning of 'bachelor' is *essentially*. For example, specifying the meaning of

'bachelor' as 'the meaning most often referred to in the recent philosophical literature on analyticity' is not appropriate precisely because it is an accidental feature of the meaning of 'bachelor' - which is in fact the one most often referred to in recent philosophical literature on analyticity - that this is the case (*ibid*: 13). In contrast, specifying the meaning of 'bachelor' as 'unmarried man' is appropriate because, unlike 'the meaning most often referred to in the recent philosophical literature on analyticity', 'unmarried man' is an essential feature of the meaning of 'bachelor.' Therefore, in giving the definition of 'bachelor', insofar as, in doing so, we are appropriately specifying its meaning, we are giving the essence of that meaning. This is supposedly consistent with the conception of words on which 'bachelor' has its meaning contingently.

Even if it is granted that definitions are indeed essence-attributions, this only applies to essence-attributions of meanings and perhaps concepts, corresponding to (on Fine's account) demonstrations of the necessity of analytic and conceptual truths. Fine's second claim that essence-attributions *simpliciter just are* definitions requires that objects and properties can be defined as well. Fine accepts that, while the idea of defining a word or concept is palatable for most philosophers, the idea of defining objects and properties is not. Against this, Fine rhetorically asks - what is so special about meanings and concepts such that we can define them and not objects? He writes,

> 'For the activities of specifying the meaning of a word and of stating what an object is are essentially the same; and hence each of them has equal right to be regarded as some form of definition.' (*ibid*: 14.)

The point here is that, as above, defining a word is appropriately specifying its meaning, and appropriately specifying its meaning just is saying what the meaning *is* (essentially); in short, defining a word is saying what the meaning of the word *is*. Fine's rhetorical question, then, is why we cannot consider our saying what (e.g.) an object *is* (essentially) as defining that object - as giving a *real definition* of that object. Assuming no reasonable answer to this question, Fine takes his second claim, that essence-attributions *just are* (real) definitions, as justified.

I now turn to whether proponents of Revelation ought to adopt this account of essence into their formulations of the thesis. As with the modal account, I will assess the

appropriateness of the real definitional account of essence based on what the account says about essence (and whether such claims have plausibility independent from Revelation) and what this means for Revelation *qua* a thesis about the essence of experience. At first glance, it might seem that the real definitional account of essence is a good fit for Revelation, given the congruence of the more traditional and commonsensical language Fine employs when discussing essence, and the language used in existing formulations of Revelation which we saw in Chapter I. For example, Fine speaks of essence in terms of a thing's *nature* (*ibid*; throughout); as we have seen, Strawson (1989: 213) too speaks of the 'essential nature' of phenomenal properties in his expression of Revelation, and Goff (2017: 107) uses the term 'complete nature' in his own formulation. Moreover, and one gets the sense that the intuitiveness of this phrasing is what motivates Fine in his claim that essence-attributions are real definitions, Fine views essence-attributions as saying 'what a thing *is*,' where this is allegedly analogous with saying what the meaning of a word or concept *is*. Here, again, we see a congruence with the language of Revelation. Goff, for example, writes that, in the case of revelatory knowledge of pain, 'I know exactly *what it is* for someone to feel that way'; this way of talking is evocative of Fine's notion of a real definition, applied to the essence of how pain feels (i.e. of the phenomenal properties which constitute the qualitative character of pain).

This congruence with Revelation, at least with the vague and brief formulations of it in the literature, is merely superficial, however. For the rest of this chapter, it will be demonstrated that the real definitional account of essence is in fact inappropriate for Revelation, due both to its independent flaws and to what those flaws mean for Revelation in particular. This demonstration will be structured around critical analysis of the two claims that Fine makes in his formulation of the real definitional account: that necessity is to be sourced in essence, and that essence-attributions are real definitions which work in the same way as giving definitions of terms. Against the first claim of Fine's real definitional account, that all necessities are to be sourced in the essences of certain entities, Penelope Mackie (2020) argues that the notion of real definition simply is not equipped to entail that essentialist-attributions have modal consequences, given that it itself is not a modal notion (as per the entire point of Fine's project); as Mackie puts it,

> 'It looks as if the account of essence in terms of real definition is intended to deliver a modal rabbit out of a non-modal hat. And I do not see how this can be done.' (*ibid*: 252.)

Mackie's particular target is the real definitional account's employment of what she calls the 'Necessity Principle' (hereafter 'NP'):

> '(NP) If being an (F) is an essential property of *x*, then being (an) F is a necessary property of *x*.' (*ibid*: 249.)

This principle, which is a part of the modal account of essence (given that, on that account, essential properties *just are* necessary properties, so F is an essential property of *x if and only if* F is a necessary property of *x*), is crucial to Fine's project of sourcing all necessities in essences. This is seen most straightforwardly in the case where a necessary truth about *x* is sourced in the essence of *x*: that necessarily I have my parents is sourced in the essential property of mine *having my parents* precisely because that essential property is also a necessary property. Likewise for the case where a necessary truth about *x* is sourced in the essence of some other entity *y*: that necessarily Socrates belongs to the set whose sole member is Socrates is sourced in the essence of that singleton precisely because the essential property of the singleton *having Socrates as a member* is also a necessary property. Generalising, finally, to cases where a necessary truth about *x* is sourced in the shared essence of a group of entities: that there are infinitely many prime numbers is (*qua* metaphysically necessary truth) sourced in the essence of *all* entities precisely because some essential properties of those entities must be necessary properties in virtue of which there being infinitely prime numbers is a necessary truth. Here we see that (NP) underlies each part of Fine's story of the 'flow' of necessity from essence; it is therefore crucial for Fine that the notion of essence that he employs, which I follow Mackie in calling 'D-essence' - that is, essence as understood via the notion of real definition and without appeal to modal notion (*ibid*) - satisfies (NP).

Mackie's argument is that there are conceptions of essence that satisfy the criteria D-essence but which do not satisfy (NP), or the particular variant of (NP) which is required by Fine, (NPD):

‘(NPD) If being (an) F is a D-essential property of *x*, then being (an) F is a necessary property of *x*,’ (*ibid*: 254)

where ‘the D-essential property of *x*’ refers to a property that is part of the D-essence of *x*, in the standard way that an essential property of *x* is understood as a property that is part of the essence of *x*. That there are conceptions of essence that satisfy the criteria for D-essence but do not satisfy (NPD) would demonstrate that the D-essence account (i.e. Fine’s account) ‘is a failure *as a basis for the project of grounding metaphysical modality on a non-modal account of essence,*’ (*ibid*; emphasis original) given how crucial (NPD) is to this project. To this end, Mackie presents two alleged examples of such conceptions of essence which could reasonably be said to satisfy the criteria for D-essence yet do not satisfy (NPD).

The first of these examples is Lockean real essences of natural kinds. On the Lockean account, real essence is defined as ‘the being of any thing, whereby it is what it is.’ (Locke 1975/1690: III.3.15.) One will immediately note that this phrasing implies that Locke conceived of essence as D-essence; in fact, Fine takes Locke as an example of a philosopher who followed the Aristotelian tradition of conceiving of essence via the notion of real definition (Fine 1994: 2). It is therefore (at least) reasonable to hold Lockean real essences to be suitable candidates for D-essences. For Locke, real essences are ‘the real internal… constitution of things, whereon their discoverable qualities depend.’ (Locke 1975/1690: III.3.17.) For example, although this was unknown during Locke’s lifetime, the real essence of water - its *internal constitution* - is ‘being H_2O.’[36] Now, the prevailing contemporary view is that ‘being H_2O’ is also a necessary property of water (e.g. as we saw, Kripke 1980), and this, taken in conjunction with Locke’s account of essence, would entail that Lockean real essences satisfy (NP), at least in the case of water. However, there are those who disagree that ‘being H_2O’ is a (metaphysically) necessary property of water. E.J. Lowe (2011), for example, an advocate for the real definitional account of essence, holds that the Lockean account of water that ‘being H_2O’ is its real essence at best entails that ‘being H_2O’ is a merely physically necessary, as opposed to metaphysically necessary (which I have simply been referring to as ‘necessary’ here), property of water. In other words, Lowe denies that the Lockean account of real essences satisfies (NP). Mackie (2020: 256) argues that it is consistent to agree with Lowe on this point while holding, as suggested just above, that

[36] Locke acknowledged that real essences, *qua* the internal constitution of things, were generally unknown (*ibid*).

Lockean real essences are D-essences - that is, essences as understood via the notion of real definition. That this is a consistent position exemplifies that the notion of D-essence does not necessitate (NPD).

Lowe, for his part, sees Lockean real essences as a radical departure from the traditional real definitional account (2011: 14), an implication of which is the mistaken (according to Lowe) view that 'being H_2O' is a metaphysically necessary property of water, so will deny that Lockean real essences are D-essences, that 'being H_2O' 'is an appropriate answer to the question "what is water?" (or "what is it to be water?"), construed as a demand for the "real definition" of the kind [*viz.* water].' (Mackie 2020: 257.) Mackie argues that the claim that 'being H_2O' is not a genuine real definition cannot be justified without appeal to modal notions (*viz.* metaphysical necessity) in an explanation of what a *genuine* real definition is, thus giving up 'on the project of providing a genuinely non-modal account of essence in terms of real definition.' (*ibid.*) This argument against Lowe does not seem fair, however. Mackie claims that Lowe does not view 'being H_2O' as figuring into a genuine real definition of water *because* it is not a metaphysically necessary property of water; if this were the case, then it is easy to see how Lowe would be forced into appealing to metaphysical necessity in order to explain what a genuine real definition is. But it is not clear that this is the case. The 'radical change of view' that Lowe identifies between the Lockean account of essence and the real definitional account is *not* that the former does not preserve the connection between essence and metaphysical necessity (although Lowe sees this as an *implication* of that change), but simply that Lockean real essences have to do with *internal constitutions*, whereas real definitions, for Lowe, have more to do with the 'macroscopic, observable' features of things (Lowe 2011: 18). So Mackie is not being fair in her claim that Lowe *must* appeal to metaphysical necessity in his account of real definition in order to justify his claims that i) 'being H_2O' is not a metaphysically necessary property of water (more generally: that the Lockean account of essences does not satisfy (NP)), and that ii) 'being H_2O' does not figure into a genuine definition of water (more generally: that Lockean essences are not D-essences). Still, even though it might be consistent for Lowe to make these claims without explaining real definitions in modal terms, it is nevertheless *also* consistent (as Mackie argues as her main point here) to agree with Lowe that the Lockean account of essence does not satisfy (NP) while also holding, contra Lowe, that 'being H_2O', along with

all of the other essences posited by the Lockean account, are D-essences, and that therefore the Lockean account does not satisfy *(NPD)* in particular.

The second example Mackie (2020: 258) offers as a consistent conception of essence which arguably satisfies the criteria for D-essence but not (NPD) is sortal concepts, as conceived by David Wiggins (1980). Sortal concepts (hereafter 'sortals') represent the property that a thing has in virtue of belonging to a certain kind. The most common kind of sortal is the substance sortal:

> 'A sortal S is a *substance sortal* if and only if, necessarily (if an individual falls under S at any time in its existence it falls under S throughout its existence).' (Mackie 2020: 258.)

On this definition, the properties represented by substance sortals are necessarily permanent (Mackie borrows this term from Parsons 2005: 9), properties that, once instantiated in some particular, cannot then be uninstantiated. Substance sortals can then be distinguished from what Wiggins (1980: 65) calls *ultimate sortals*:

> 'A sortal S is an *ultimate sortal* if and only if S is *the most general* sortal corresponding to some principle of individuation (or criterion of identity).' (Mackie 2020: 258.)

The idea behind substance sortals is that in cases where substance sortals are individuated by the same criterion of identity (e.g. 'cat' and 'dog'), there is some more general sortal that corresponds to this criterion (e.g. 'carnivoran'); ultimate sortals are maximally general sortals which correspond to some criterion of identity (e.g. 'mammal').[37] For Wiggins, only these ultimate sortals are what Mackie calls 'necessary sortals,' where,

> 'A sortal S is a *necessary sortal* if and only if the thing that falls under S could not have existed without falling under S.' (Mackie 2020: 259.)

[37] Wiggins is frustratingly unclear in his writings on ultimate sortals, giving no actual examples. He claims that not all animals share a principle of individuation (1980: 122-123), presumably implying that he does not see 'animal' as an ultimate sortal, hence the example I offer being 'mammal' - something more general than 'dog' but less general than 'animal'. The obscurity does not matter much for Mackie's purposes here, the point is, as I explain below, that it is consistent to distinguish between two kinds of sortal, one that is necessary and one that is merely necessarily permanent.

It is important to note the difference between this definition and that of substance sortals above. As above, the properties represented by substance sortals are necessarily permanent. The difference between necessarily permanent properties and necessary properties - i.e. the properties represented by necessary sortals, as defined above - is as follows. A property F is *necessarily permanent* if and only if that *x* if F entails that *x* is F so long as *x* exists. Crucially, this does not entail that *x could not have been not F*; this further entailment only comes with F being a *necessary* property. In terms of possible worlds, if F is a necessarily permanent property, that *x* is F at world *w* does not entail that *x* is F is any world other than *w*, only that *x* if F throughout the existence of *x at world w* (*ibid*); if F is a necessary property, on the other hand, that *x* if F in *any* world entails that *x* is F in *all* worlds. As defined above, then, that S is a substance sortal does not entail that it is a necessary sortal, and so it is consistent to hold, as Wiggins does, that substance sortals are not (necessarily) necessary sortals.

Mackie further observes, in parallel with her first argument regarding Lockean real essences, that the properties represented by substance sortals are good candidates for being D-essential properties. Furthermore,

> 'On the assumption that *horse* is a substance sortal, to say, of a horse, that it is a horse (as opposed, say, to saying that it is brown, or neighing, or in the stable, or an Ascot winner) appears to be an eminently plausible answer to the (Aristotelian) questions "what is it?", "what is it to be the thing that it is?," even if we think that the horse could have existed without being a horse. Substance sortals seem to be admirable candidates for the role of D-essences, regardless of whether they are necessary sortals.' (*ibid*: 260.)

In other words, it is consistent to hold that the properties represented by substance sortals are D-essential properties while also holding that these properties are not (necessarily) necessary properties; or, to hold that substance sortals are 'D-essential sortals' while also holding that substance sortals are not (necessarily) necessary sortals. As Mackie observes, this was in fact Wiggins's own view (see Wiggins 1980: chs. 2&3). As with Mackie's first argument, this again demonstrates that the notion of D-essence does not entail that D-essential properties are necessary properties, that the notion does not necessitate (NPD).

Lowe (2007: 765) disagrees with Wiggins and Mackie that it is consistent to hold that substance sortals are not necessarily necessary sortals, arguing that substance sortal S's being a necessary sortal is the only way to explain the fact that (as per the above definition of substance sortals) if *x* falls under S then *x* cannot cease to fall under *x* without ceasing to exist. This is because, avers Lowe, an inquiry as to which substance sortals *x* falls under is an inquiry as to the '*essence* or *nature*' of *x* (*ibid*; emphasis original). Here Lowe is claiming that substance sortals are representative of essences, or, more precisely, that the properties represented by substance sortals are essential properties. From this explanatory claim that the property represented by S is an essential property of *x* (insofar as *x* falls under S), it follows, according to Lowe, that the property is also a necessary property of *x*, and therefore that S is also a necessary sortal. In other words, S's being a substance sortal under which *x* falls can only be explained by S representing some essential property of *x*, and S's representing some essential property of *x* entails that S is a necessary sortal. It follows from this that it is not a consistent view to hold that substance sortals are D-essential sortals but not necessary sortals, a view that would not satisfy (NPD).

As Mackie points out, however, this response begs the question of whether a D-essential property of *x* is also a necessary property of *x*. To see this, note that Lowe is equating the question of which substance sortals *x* falls under with the question regarding the essence *or nature* of *x*. The equivocation between essence and nature is a staple of the real definitional account of essence (and is littered throughout Fine 1994, in particular), and so it would be fair to take Lowe - *qua* advocate of that account - to be saying that in asking which substance sortals *x* falls under we are inquiring as to the *D-essence* of *x*, that the property represented by S is a *D-essential* property of *x*. But now the question-begging is clear, located at the point at which Lowe moves from that claim to the claim that (therefore) the property represented by S is a necessary property, given that this move - from something's being a D-essential property to its being a necessary property - is precisely what is at issue here. Mackie is arguing that it is consistent to hold that substance sortals are not necessarily necessary sortals *even if* they represent D-essential properties; Lowe is not entitled to argue that it is in fact *inconsistent* to hold that substance sortals are not necessarily necessary sortals *because* they represent D-essential properties and *therefore must* be necessary sortals. Mackie's second argument, therefore, stands, and Fine's claim that, on the real definitional account, necessity can be sourced in essence, is false.

Here, one, while in agreement with Mackie's points and subsequent conclusion that Fine cannot ground necessity in essence with real definitions, might wonder why this should worry the proponent of Revelation who nevertheless wishes to adopt the Finean real definitional account into her own preferred formulation of the thesis. The general project of accounting for modality in terms of essence, of which Fine's first claim of grounding necessity in D-essence is an expression, is, as above, *essentialism* about modality. While this is a popular view about modality, it is by no means free of detractors. Indeed, Mackie, in arguing that D-essence cannot entail necessity in the way that Fine wants, does not herself hold an alternative essentialist view about modality. That is, the target of her argument is not just *Fine*'s essentialism, but essentialism *in general*.[38] The question for our purposes, then, is, why must the proponent of Revelation be an essentialist about modality, why should it matter to them that Fine's essentialism about modality fails? While a full argument for essentialism about modality lies outside the scope of the present thesis, I will here suggest why, regardless of the independent merits and flaws of essentialism about modality in general, our proponent of Revelation ought to be an essentialist about modality. It is worth conceding first, however, that the argument from Revelation to physicalism, formulated by Lewis (as the incompatibility of the two theses), does not appear to require a commitment to essentialism about modality. According to that argument: Revelation entails that the full set of properties which are part of the essence of an experience are revealed in introspection, physical properties are not among this set, therefore the essence of that experience does not involve any physical properties; clearly, essentialism about modality is not required for this argument, given the lack of modal notions involved.

Nevertheless, I suggest that the proponent of Revelation ought to be an essentialist about modality, at least with respect to their formulation of Revelation, because the necessity principle (NP) is implicit in the general anti-physicalist dialectic of which Revelation is a part. For example, the common sense which Kripke appeals to in demarcating the rigid designators from the non-rigid designators implicitly involves a commitment to essentialism: the reason why 'pain,' for example, is a rigid designator is because being a pain is part of the *essence* of pain, and *as a result* it is *absurd* to suppose that pain *could have been* something

[38] See also Leech (2020).

other what it is, in essence.[39] This appears to be an implicit commitment to (NP): pain is necessarily pain *because* what it is, in essence, is pain. Recall that, in §1, I argued that it is this same common sense which makes the Kripkean principle, that identity statements involving two rigid designators are necessarily true if true at all, similarly commonsensical. Furthermore, I argued, from this claim that that Kripkean principle is commonsensical, and also from the claim that mind-brain identity statements do not involve descriptive content, to the claim that the intuition that mind-brain identity statements are possibly false collapses into the intuition that such statements are *actually* false. In other words, in arguing that the intuition of possible distinctness collapses into the intuition of (actual) distinctness, I appealed to the fact that the necessity-of-identity principle is commonsensical which, if that common sense, as I have just suggested, is implicitly committed to (NP), implies that my argument that the intuition of possible distinctness just is the intuition of actual distinctness is likewise so committed. Recall also that, §3, this claim that the two intuitions of distinctness are the same was crucial to my argument that Revelation implies that there should not be (a widespread intuition of possible distinctness, which in turn implies that there should not be) a widespread intuition of distinctness. To the extent, therefore, that our proponent of Revelation wishes to plug the hole in the '*p* seems false, so *p* is false'-style arguments we discussed in Chapter I, she ought to adopt an account of essence which implies (NP).

The proponent of Revelation is nevertheless not, as I conceded, *required* to be an essentialist about modality, and Revelation will work against physicalism in the more straightforward way regardless. She is therefore *entitled* to drop the Finean commitment to essentialism while continuing to adopt the rest of the real definitional account into her formulation of Revelation. Regardless of whether Revelation needs the alleged modal rabbits from the real definitional account, however, the account suffers from a much more fundamental problem, one that can't be separated from the account itself - namely, Fine's second claim that objects and properties can be defined in the same way as meanings and concepts, where definitions here are considered to be attributions of essence. Recall that, in order to justify this claim, Fine first argues that definitions in the usual sense, that is, definitions of words, insofar as they function to specify the meaning of words, where meaning is a word's essence, are essence-attributions; then, to extend this to the essence-attributions of

[39] Recall, Kripke writes, 'if something is a pain it is essentially so, and it seems absurd to suppose that pain could have been some phenomenon other than the one it is.' (Kripke 1980: 149.)

objects and properties, Fine simply poses the rhetorical question: what is so special about words that we can define them but not objects and properties? This question is posed right at the end of his paper 'Essence and modality,' (1994) and Fine does not go onto explaining this notion of a 'real definition' - that is, the definition of an object or property. The issue, however, is that we do in fact have quite good *prima facie* reason to treat words and objects/properties differently in this way. On the one hand, words and concepts are linguistic objects whose 'essences,' or meanings, in their case, are fixed by convention or specification. It is for this precise reason that words can be defined. The essences of objects and properties, on the other hand, are *out there*, objective and mind-independent, they are not fixed by convention or specification. Such essences are *discovered*, not simply specified. It is therefore unclear as to how we may nevertheless *define* objects and properties *in the same way* as we define words, given that definition is usually, again, a matter of convention or specification. Rather than pose the rhetorical question, What is so special about words that we can define them but not objects?, assume no reasonable answer, and take the notion of real definition to be thereby clarified, the onus is on Fine to give a detailed account as to how, given that words and objects *are* so different in the above way, they can nevertheless be defined *in the same way*.

Without such an account, the notion of real definition, and by extension D-essence, is left obscure, and this makes it difficult to discern exactly what any given essence-attribution is referring to. Proponents of the real definitional account must fall back on the locution 'what a thing is,' but, if the 'is' here is understood in this obscure definitional sense, this does not help in specifying what the D-essence of any given object or property is. As Mackie's critique of the real definitional account demonstrated, if the criterion for being a D-essence is simply something which would make a suitable answer to the question 'what is it?', then there are all sorts of candidates for what D-essence could be: for example, it was seen to be equally plausible to hold that, with Locke, that D-essences are 'internal constitution' properties, like 'being H_2O,' as it was to hold, with Lowe, that D-essences are closer to appearance properties, like 'watery stuff,' or to hold that D-essences are more like properties represented by substance sortals. In fact, it was this minimal criterion which made the real definitional account vulnerable to Mackie's critique in the first place: all that Mackie had to do was find some sort of property that plausibly fit the incredibly minimal criterion of being a suitable

answer to the question 'what is it?', and show that the instantiation of that sort of property does not (necessarily) have modal consequences.

This obscurity is not something that our proponent of Revelation ought to adopt into her formulation of the thesis. Firstly, with no clear idea as to how objects and properties are supposed to be 'defined,' it is difficult to understand exactly what Liu means in her formulation of Revelation as the thesis that, in introspecting an experience, we come to know 'Q is X' where the predicate 'X' *captures* the essence of the phenomenal property Q, which, in other words, means that 'X' *defines* Q.[40] Secondly, and perhaps more importantly, the vagueness of what sort of property (internal constitution, accidental-appearance, etc.) D-essence is supposed to be, on the real definitional account, allows that Revelation might not contradict physicalism in the straightforward Lewisian sense after all. For example, in parallel to Lowe's holding that the D-essence of water is not H_2O, it might be said that the D-essence of pain is not C-fibres firing. It may well still be that C-fibres firing is the Lockean internal constitution of pain, and this might be enough for the physicalist, who would have to say that 'pain is C-fibres firing' is not a statement of (D)-essence after all. In this case, the fact that it is not revealed in introspecting a pain experience that the (D-)essence of pain is C-fibres does not contradict physicalism, and in particular the claim 'pain is C-fibers firing,' even if Revelation is true. Given this particular implication of the obscurity of the real definitional account of essence, but also that obscurity in general, the proponent of Revelation ought not adopt that account of essence into her formulation of the thesis.

* * *

In this chapter, I have examined two accounts of essence and how well those accounts would fit into formulations of Revelation, finding both to be ill-suited for adoption into such formulations. I endorsed Fine's critiques of the modal account of essence, agreeing that it had unwanted implications about what sorts of properties are included in a thing's essence; in particular, I demonstrated that formulations of Revelation which adopted the modal account would have inordinately strong implications as to the would-be revelatory knowledge gained

[40] This obscurity further muddies what are already claggy waters with regards to what the essence of Q is supposed to be, given that, as Liu rightly observes, whatever the essence of Q is, it is already hard to put into words (Liu 2019: 232).

through introspection. Next, against the alternative, real definitional account favoured by Liu in her own formulation of Revelation, I levelled two critiques. First, I endorsed Mackie's argument that the real definitional account fails to ensure that essence-attributions have modal implications; here I argued that this failing of the real definitional account, if that account were to be applied to Revelation, would block the argument I made in Chapter I that Revelation entails that there should not be a widespread intuition of distinctness. Second, I argued that the notion of real definition is obscure, and that this makes the real definitional account of essence too vague on what exactly essence is; here I argued that this vagueness, inherited by formulations of Revelation that adopted that account of essence, would weaken Revelation against physicalism, such that, on certain understandings of 'essence' which the real definitional account allows, the truth of Revelation would be consistent with the truth of physicalism. In light of the negative conclusions of this chapter, proponents of Revelation will have to look elsewhere for an account of essence.[41]

[41] A promising candidate is the identity account of essence, due to Correira and Skiles (e.g. 2019) which (i) has better prospects of securing (NP) (although see Leech 2020 for an argument against these prospects), (ii) gives a precise analysis of what essence actually *is*, and (iii) avoids the undesirable implications of the modal account.

Concluding remarks

Over the course of this thesis, I have offered two contributions to the literature surrounding Revelation and its place in the anti-physicalist project. The first is the argument that Revelation entails that there should not be a widespread intuition of distinctness; applying this to various anti-physicalist arguments - due, respectively, to Kripke, Levine, and Strawson - which, I have shown, all make appeals to that intuition, means that Revelation is able to plug the hole that is common to all of them, namely the gap from '*p* seems false' to '*p* is false.' My second contribution is my argument that the real definitional account of essence, at least as Fine formulates it in 'Essence and modality' (1994), is not the simple alternative to the modal account of essence which proponents of Revelation, such as Liu, might be tempted to think. Future dialectic surrounding Revelation ought to (i) find a more appropriate account of essence for proponents to adopt, and (ii), perhaps more crucially, provide further substantive argument for the *truth* of Revelation, given its importance, as we have seen over the course of the present thesis, to the anti-physicalist project.

Bibliography

Brewer, B. 2019. ‘Visual experience and the three R's’ in J. Knowles & T. Raleigh (eds.) *Acquaintance: New Essays*. Oxford: OUP.: 277-292.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: OUP.

Chalmers, D. 2010. *The Character of Consciousness*. Oxford: OUP.

Coleman, S. 2015. ‘Neuro-cosmology’ in P. Coates and S. Coleman (eds.) *Phenomenal Qualities: Sense, Perception, and Consciousness*. Oxford: OUP.: 66-102.

Coleman, S. 2016. ‘Panpsychism and neutral monism: how to make up one’s mind’ in G. Bruntrup, & L. Jaskolla (eds.) *Panpsychism: Contemporary Perspectives*. Oxford: OUP.: 249-282.

Coleman, S. 2019. ‘Natural acquaintance’ in J. Knowles & T. Raleigh (eds.) *Acquaintance: New Essays*. Oxford: OUP.: 51-74.

Correira & Skiles. 2019. ‘Grounding, essence, and identity.’ *Philosophical and Phenomenological Research* 98(3): 642-670.

Fine, K. 1994. ‘Essence and modality.’ *Philosophical Perspectives* 8: 1-16.

Goff, P. 2017. *Consciousness and Fundamental Reality*. Oxford: OUP.

Hempel, C. 1980. ‘Comments on Goodman’s *Ways of Worldmaking*.’ *Synthese* 45: 193-200.

Johnston, M. 1992. ‘How to speak of colors.’ *Philosophical Studies* 68(3): 221-269.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Leech, J. 2020. ‘From essence to necessity via identity.’ *Mind* 130(519): 887-908.

Levine, J. 1983. ‘Materialism and qualia: the explanatory gap.’ *Pacific Philosophical Quarterly* 64: 354.361.

Levine, J. 2001. *Purple Haze.* Oxford: OUP.

Lewis, D. 1995. ‘Should a materialist believe in qualia?’ *Australasian Journal of Philosophy* 73(1): 140-144.

Loar, B. 1990. ‘Phenomenal states.’ *Philosophical Perspectives* 4: 81-108.

Liu, M. 2019. ‘Phenomenal experience and the thesis of revelation’ in D. Shottenkirk, M. Curado, & S.S. Gouveia (eds.) *Perception, Cognition and Aesthetics*. New York, NY: Routledge.: 227-251.

Liu, M. 2021. ‘Revelation and the intuition of dualism.’ *Synthese* 199: 11491-11515.

Locke, J. 1975. *An Essay Concerning Human Understanding*, P. Nidditch (ed.). Oxford: OUP. Originally published in 1690.

Lowe, E.J. 2007. 'Sortals and the individuation of objects.' *Mind & Language* 22(5): 514-533.

Lowe, E.J. 2011. 'Locke on real essence and water as a natural kind: a qualified defence.' *Proceedings of the Aristotelian Society, Supplementary Volumes* 85: 1-19.

Mackie, P. 2020. 'Can metaphysical modality be based on essence?' in M. Dumitru (ed.) *Metaphysics, Meaning, and Modality: Themes from Kit Fine*. Oxford: OUP.: 247-264.

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

Nide-Rümelin, M. 2007. 'Grasping phenomenal properties' in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: OUP.: 307-336.

Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.

Papineau, D. 2007. 'Kripke's proof is *ad hominem* not two-dimensional.' *Philosophical Perspectives* 21: 475-494.

Papineau, D. 2021. *The Metaphysics of Sensory Experience*. Oxford: OUP.

Parsons, J. 2005. 'I am not now, nor have I ever been, a turnip.' *Australasian Journal of Philosophy* 83: 1-14.

Raleigh, T. 2019. 'The recent renaissance of acquaintance' in J. Knowles & T. Raleigh (eds.) *Acquaintance: New Essays*. Oxford: OUP.: 1-30.

Russell, B. 2001. *The Problems of Philosophy*. Oxford: OUP. First published 1912.

Spurrett, D. & Papineau, D. 1999. 'A note on the 'completeness' of physics.' *Analysis* 59(1): 25–29.

Strawson, G. 1989. 'Red and 'Red'.' *Synthese* 78(2): 193-232.

Strawson, G. 1994. *Mental Reality.* Cambridge, MA: MIT Press.

Strawson, G. 2006. 'Realistic monism: why physicalism entails panpsychism.' *Journal of Consciousness Studies* 10-11: 3-31.

Stoljar, D. 2008. 'The argument from revelation' in D. Braddon-Mitchell & R. Nola (eds.) *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press.: 113-138.

Wiggins, D. 1980. *Sameness and Substance*. Oxford: Blackwell.