

**Mary and the Philosophical Goose Chase:
*Physicalism and the Knowledge Argument***

Introduction

0.1 In his paper ‘Epiphenomenal Qualia’, Frank Jackson introduces his famous thought experiment as follows:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room *via* a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like ‘red’, ‘blue’ and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces *via* the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence ‘The sky is blue’...

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. *Ergo* there is more to have than that, and physicalism is false. (1982, p.130)

This version of the Knowledge Argument has proven a real pain for physicalist philosophers (eventually including Jackson) since it first appeared in the early 1980s. Much debate and no doubt countless hours of solitary philosophical contemplation have been spent on defending physicalism against the imprisoned colour scientist and as a result many different responses have appeared demonstrating varying levels of ingenuity and success.

There are several different points at which the argument can be disputed and philosophers have dutifully fallen into the different camps, each offering their own version of the principal objections and showing why their version should be preferred to all the others. In what follows I shall endeavour to show that the vast majority of philosophers seeking a decisive refutation to Jackson’s argument have simply been looking in the wrong place and have thus contributed to the enduring ‘wild goose chase’ that has wasted an awful lot of philosophical time; time that could perhaps have been put to better use.

0.2 The format for this work will be as follows. In section 1 I shall present Jackson’s version of the Knowledge Argument, after a brief nod towards its previous incarnations, in order to set the scene. I shall then discuss the varying points at which objections can be made and briefly survey the most popular approaches.

In section 2 I shall consider in more detail those objections that focus on the idea that the new knowledge some say Mary ‘must surely’ gain when she first sees a colour is non-factual, and thus no threat to the truth of physicalism. Focusing on both the Ability Hypothesis and the related Acquaintance Hypothesis, I will demonstrate that where the former seems to require too much of Mary in order to know what an experience is like, the latter isn’t clear on what is required at all. Neither, therefore, is capable of refuting Jackson’s argument.

Section 3 will delve into the popular, but often complex, objection that Mary does indeed gain new factual knowledge, but this is no threat to physicalism. This is either because what Mary actually gains is new knowledge of an old fact she already knew, or learns a new phenomenal fact which refers directly to a physical property. I will show that both of the approaches considered should be dismissed. Partly due to confusion over modes of presentation and phenomenal facts, but more importantly because an approach focusing on philosophy of language completely misses the point of what the debate should be about and actually plays right into the hands of the dualists by unnecessarily separating the physical and the phenomenal.

In section 4 I shall take a bit of a step back from the mainstream and consider Daniel Dennett’s die-hard materialist approach, which claims that Mary can know everything there is to know whilst still confined in her black and white environment and without ever experiencing a colour. Here I will show that although this approach is not (surprisingly) as silly as it sounds, it is terribly confused and thus threatens to entail a blanket denial of experiential reality which – though some accept it – *is* just as silly as it sounds.

I shall conclude by conceding that there may in fact be no apparent way to conclusively refute Jackson’s argument. However, I will also demonstrate that this admission need not trouble *real* physicalists as Jackson’s thought experiment only succeeds in refuting the kind of physicalism that no one should support.

1. All aboard the Knowledge Argument train.

1.1 The Knowledge Argument has been around for many years, appearing in different works in different formulations. There are many reasons why Jackson’s version of it has attracted so much attention, and I will come to these in due course. First of all it would be well to consider a couple of earlier incarnations of the argument.

The Knowledge Argument is designed to show the evident gap between our physical descriptions of the world and our actual experiences, proving that complete physical knowledge isn’t sufficient for complete knowledge of phenomenal states and therefore our experiences must involve non-physical properties. Versions of the argument have been kicking around since Locke (and probably before), but one distinct formulation came with C.D Broad’s “mathematical archangel” in 1925. Broad’s creation is a logically omniscient being who thus knows all the physical truths relating to various chemical compounds. Such knowledge, however, maintains Broad, would not mean the archangel has the

phenomenal knowledge concerning, for example, the “peculiar smell of ammonia”; thus he infers that physicalism is false.

He [the archangel] would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself. (1925, p.71)

Another popular argument comes from Thomas Nagel’s seminal paper published in 1974, “What is it like to be a Bat?” – a paper which also launched the concept of *what it is like* to have a particular experience into philosophical terminology and begun a long tradition of interpretation and re-interpretation which is in fact responsible for much of the chaos that surrounds discussions of phenomenal consciousness today. (In fact, it will become apparent that confusion over this phrase and its exact meaning has contributed greatly to the problems involved in the Knowledge Argument. Once it is defined properly, many uncertainties and misunderstandings could most likely be cleared up.) Nagel’s point is simple: even if we knew everything there is to know from an *objective* perspective about a bat’s sonar system, we still would not know what it is like to use such a system in order to perceive an object.

Though these arguments undoubtedly illustrate a notion which may seem latent in human thinking, referred to by Nagasawa and Stoljar (2004) as the knowledge intuition – basically that no amount of knowledge of a physical sort is going by itself to suffice for knowledge of a phenomenal sort – they are not specifically targeted against physicalism in the way Jackson’s formulation is. Nagel’s arguments that humans cannot possibly know what it is like to be a bat raise as many questions about the limits of human imagination as they do about the truth of physicalism, and whilst Broad’s may not have this problem it is perhaps open to the physicalist just to deny that the archangel would lack the relevant phenomenological knowledge. Broad makes the basic point, but what Jackson’s argument does is provide a situation in which it is much harder to deny that the subject learns anything new.

The Mary formulation is more powerful an argument than its predecessors not just because of its explicit focus on the falsity of physicalism, but also because of its superior illustration of the knowledge intuition and therefore the conclusion it entails. Unlike Broad’s example, denying that Mary lacks the knowledge of what it is like to see a red rose whilst still confined in her room leads to specific problems, as she does clearly seem to learn this on her escape; thus logic dictates that if she learns it on coming out of her room she did not know it before. With regard to Nagel’s arguments, it could be said that physicalism clearly does not entail that we *as humans* must be able to imagine the particular experiences a bat has; Jackson’s argument obviously avoids this problem.

1.2 As I mentioned in the beginning there are many different points in the argument at which objections can be made to stop it in its tracks. In order to see the different parts, it is helpful to lay the argument out as follows:

1) Mary has all the *physical* information¹ about colour vision before her release

2) However, there is some information about human colour vision that she does not have before her release (namely – she does not know what it is like to experience colour)

Therefore

3) Not all information is physical information.

It seems that the majority of physicalist responses have ignored premise one, happily accepting that their position means it is, at least, *logically* possible that Mary could come to know everything physical there is to know about colour vision from within her black and white prison². Instead they have focused their attention on the questions of, firstly, whether or not Mary learns anything new on her release, and secondly, if she does (the more popular road), just what kind of knowledge could she gain that would be compatible with physicalism. Going through the steps of the argument then throws up a number of questions which sorts out the responses. The key ones can be stated as follows³:

Q1. Does Mary actually learn anything new or gain any knowledge when she leaves her room and first has an experience of colour?

Q2. If Mary does gain new knowledge, is it *factual/propositional* in nature?

Q3. If it seems Mary does gain propositional knowledge, does she gain knowledge of *new* propositions or just come to know *old* propositions in a new way?

A ‘no’ answer to **Q1** would perhaps seem appealing to the most die hard eliminative materialist or scientist, but (as will be shown later) it is hard to maintain. One approach is to say ‘no’ to **Q2** and claim that Mary’s new knowledge is not propositional in nature; rather it is either the gaining of certain abilities, or only acquaintance knowledge. These arguments have some appeal and will be considered in the next chapter, though the versions presented ultimately fail. An acceptance that Mary’s learning does involve propositions then takes us on to **Q3** and a number of philosophers have chosen to go down this route, arguing that Mary comes to know old facts in a new way, or that she learns new phenomenal facts made true by physical properties.

As I am not able to deal with every physicalist response in what follows, I shall restrict myself to what I consider to be the main proponents of the particular views⁴.

2. Getting off at Q2: Abilities and Acquaintance.

2.1 Leaving aside a negative response to **Q1** for the moment (an unpopular approach), let us consider the views of those who try to get off the Knowledge Argument train by answering ‘no’ to **Q2**. These philosophers can claim that though Mary *does* learn

something new when she is released from colourless captivity, she does not gain any new propositional knowledge. Thus the incompleteness of Mary's pre-release knowledge in no way entails a rejection of physicalism. In this section I propose to consider two such approaches, known as the Ability Hypothesis and the Acquaintance Hypothesis. I shall demonstrate that of the two the former can be shown to be obviously flawed and the latter too ill-defined to mount a serious challenge against the Knowledge Argument.

2.2 The Ability Hypothesis says that the knowledge Mary gains is not knowledge of any new facts, but rather a kind of non-informational *know-how*, which consists in the abilities to remember, imagine, and recognise colour experiences. One key proponent of this view is David Lewis (1988) who argues that knowing what an experience is like just *is* the possession of these abilities⁵. Thus what Mary acquires is more mental skills than new facts.

Lewis agrees with Jackson that experience is undoubtedly the best teacher and when Mary steps out of her black and white environment and sees a red rose, or green grass she does indeed learn what it is like to experience red or green. No one who has never tasted vegemite knows what it is like to taste vegemite and the only way they can learn what it is like is to have that experience⁶. Once they have done so they can then remember the taste; can imagine what it might be like to taste other things with the same flavour (like vegemite ice cream); and can recognise the taste when it comes again:

These abilities to remember and imagine and recognise are abilities you cannot gain (unless by super-neurosurgery, or by magic) except by tasting Vegemite and learning what it's like. You can't get them by taking lessons on the physics or the parapsychology of the experience, or even by taking comprehensive lessons that cover the whole of physics and parapsychology⁷. The ability hypothesis says that knowing what an experience is like just *is* the possession of these abilities to remember, imagine, and recognise. It isn't the possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities aren't actualised. It isn't knowing-that. It's knowing-how. Therefore it should be no surprise that lessons won't teach you what an experience is like. Lessons impart information; ability is something else. Knowledge-that does not automatically provide know-how. (1988, p.99-100)

When Mary leaves her black and white environment and sees a red rose she learns what it is like to have that experience: meaning that she gains the abilities to remember what it is like to see red; she can now imagine what it is like and apply it to other situations (she can, say, imagine what a red car looks like even though she hasn't yet encountered one); she can also recognise the experience when it comes again, recognise other red things etc. The phrase 'knowing what it is like' just means to have these abilities: to have the knowledge-how.

2.3 There have been a number of objections levelled at the Ability Hypothesis, mostly concerned to demonstrate that having the abilities Lewis discusses is neither necessary nor sufficient for knowing what it is like to have a particular experience. The way this is demonstrated is usually to show how you can have the one without the other, so the *is* that Lewis uses cannot be the *is* of identity, as he intends it to be.

Earl Conee (1994) presents the example of a person who, for whatever reason, has no visual imagination, and will be unable to imagine, say, an experience of red. Lewis' view would seem to entail that such a person would not know what it is like to experience red, even *whilst looking* at a red rose. This seems absurd, so we might conclude that knowing what it is like is not one and the same thing as the possession of the Lewisian abilities.

Torin Alter (2001) uses a similar example of mental impairment to show further problems with the Ability Hypothesis. He argues specifically that in addition to separating the abilities from knowing what an experience is like, we can even separate ability from know-how. His example goes as follows:

Suppose Mary doesn't leave the black and white room. Instead, a trusted friend slips red paper under the door and informs her that the paper is red. Having seen the red paper, she knows how to imagine red experiences. Then a brain injury affects her general imaginative ability; she can't visualise anything. Does she lose her knowledge of how to imagine red experiences? Not necessarily, as we might discover if she recovers her ability to imagine as the effects of the injury recede. She recovers the ability to imagine *red* experiences, not blue ones, without leaving the room. Evidently, she retains her know-how without retaining her ability. So, her know-how can't be identified with the corresponding ability. Similar reasoning applies to knowing how to remember and recognize the experiences. So, the know-how relevant to the Ability Hypothesis can't be identified with the corresponding abilities. (2001, p.235)

Alter is arguing that here Mary loses her ability to imagine red experiences but retains the knowledge of *how* to do so, as she recovers it upon recovering from her injury. She retains the know-how in the absence of the ability, thus they cannot be one and the same thing.

What can be said here? Perhaps it is possible for Lewis to claim simply that what is retained is not knowledge or know-how, as to qualify as that would require a certain sort of access to it. Yet still it appears that something is retained in spite of the loss of ability, therefore it is something which is not reducible to the ability. Perhaps Lewis then must just claim that what Mary retains is in fact the ability, only exercising it is not possible during the period of her injury. The problem then is that whilst suffering the braindamage it could be argued that Mary does not know what it is like to experience red, despite retaining the ability, so either way it seems difficult to maintain that they are one and the same thing.

2.4 If we find these examples problematic because we don't know the exact details of what a loss of visual imagination would entail, we can consider a slightly different example, given by Michael Tye (2000), which again draws into question whether knowing what an experience is like is identical to possessing the corresponding abilities.

Human sensory experience is enormously rich and, in the case of colour experience, we have no memory impressions for most hues. When Mary sees a red rose she acquires the relevant Lewisian abilities: she can now recognise red things, remember the experience of red, can imagine what it is for something to be red. But this is not all. The rose will be a particular hue – say red17 – and while she can know what it is like to experience red17 at

the time she is looking at it, she will not be able to save detailed enough information to retain the Lewisian abilities with regards to red17. If, then, Mary does not gain any of these abilities with regards to red17, and 'knowing what it's like' just *is* to have these abilities, we seem to be obliged to say that Mary does not know what it is like to experience red17 even when she is looking at it – so we have the same issue as before, in the case in which Mary lacked any visual imagination, but without the difficulties of knowing exactly what that would entail.

What this example does, then, is show that the Ability Hypothesis creates a distinction between knowing what it is like to see red and knowing what it is like to see red17. Arguably we would want to say that Mary knows both what it is like to experience red and red17 (the latter at least whilst she is looking at it), but if this is the case then knowing what an experience is like cannot literally be one and the same thing as the possession of the corresponding abilities, because they are not present with regard to red17. Lewis' definition seems to require too much of a person for it to be true to say that they know what an experience is like.

2.5 Is it false, then, to maintain that knowing what an experience is like just is the possession of abilities to remember, imagine and recognise? Let us see if we can save the Ability Hypothesis by considering the following example.

Mary is indeed afflicted with a total lack of visual imagination. One day she is allowed to leave her black and white room and go next door into a room that is completely green. As discussed, it seems difficult to say that whilst in the green room Mary does not know what it is like to experience green, indicating that the Ability Hypothesis is false. However, maybe it could be argued that the problem with visualising colours happens only in their absence, and whilst actually in the presence of the particular external stimuli Mary *would* be able to imagine, remember and recognise the same colour experience. If this were the case then we could reconnect the abilities and knowing what an experience is like. Whilst in the green room Mary would have all the abilities and thus know what the experience is like, then after returning to the black and white environment she would lose the abilities and thus no longer know what it is like.

As I have no real idea of what a lack of visual imagination would entail, it is possible this attempt to save Lewis' theory could be dismissed on the grounds that it is not possible for someone with such a condition to have the abilities in the situation I have described. Perhaps, though, it can be applied instead to Tye's objection involving the different hues.

This time Mary has no visual imagination problems when she steps into the completely green room, but she is confronted with green17. Again, it is difficult to say Mary does not know what it is like to experience green17 whilst in this room, but when she returns to her black and white environment she will only have the Lewisian abilities with regards to 'green' in general. Is it possible to argue that whilst in direct contact with that particular shade of green Mary *does* possess all the required abilities (with regard to green17), only she loses them on her return to the black and white environment (retaining only abilities

with regard to 'green' in general), and thus no longer knows what such an experience is like?

I doubt that anyone would find either of these suggestions particularly appealing. The abilities one gains on having a particular experience of colour are unlikely to ever be so specific, even when confronted with a particular shade, and the connection between Lewis' abilities and knowing what an experience is like remains far weaker than he is suggesting. Combined with Alter's objection that we cannot even connect know-how and ability in the way Lewis wants it, the Ability Hypothesis is unlikely to rid us of Jackson's Mary.

Lewis just seems to be asking far too much of Mary in order for her to know what an experience is like, suggesting an alternate approach to the problem – which is where the Acquaintance Hypothesis comes in.

2.6 The Acquaintance Hypothesis still rests on the idea that Mary does not learn any new facts after she leaves the room, despite coming to know what an experience is like, but does not require her to possess the abilities just discussed. Rather, all Mary needs to do is notice the experience as it is undergone.

Earl Conee (1994) presents a version of this view and claims that what Mary gains is neither factual knowledge, nor ability knowledge but a different type of knowledge altogether, known as *acquaintance knowledge*. There are two main problems with Conee's presentation of this position. The first, highlighted by Alter (1998), shows that the analogies Conee draws on when presenting his theory fail to provide the intuitive support expected of them, and the second arises due to Conee's stipulation that in order to be acquainted with an experience one must 'notice' it as it is undergone.

Acquaintance knowledge, as Conee presents it, does not involve knowledge of facts or possession of abilities but rather a sort of direct epistemic relation to something (such as a phenomenal property). He compares it to having knowledge of a city only after you have visited it, or of a person only after you have met them. You may know all the facts about the city or the person prior to direct contact, but you can only gain acquaintance knowledge once you have visited the actual place, or met the actual person, and only then can you come to know what they are like⁸. So with regard to colour experiences, it is true that Mary does not know what it is like to, say, experience red whilst confined in her room; she comes to know such an experience once she escapes and becomes directly acquainted with the experience and its phenomenal quality.

Conee jumps on Jackson's notion that Mary would be ignorant of a property or quality related to colour experience and maintains that to be ignorant of a property is not to fail to know a fact, nor is it to lack an ability, but just to lack acquaintance with the property. Thus, to come to know a property is to become acquainted with the property in the same way as to come to know a city is to be acquainted with the city. Therefore:

A simple acquaintance hypothesis about what Mary learns is that learning what an experience is like is identical to becoming acquainted with the experience. When Mary first sees red ripe tomatoes, she learns what it is like to see something red. It is also true of this episode that it is the first time that she undergoes an experience with the phenomenal quality that ordinarily results from seeing something red, phenomenal redness. (1994, p.202)

So if the phenomenal quality is a physical property of the experience, then Mary with her exhaustive knowledge will know that experiences have this property, but she will not know what it is like to see red. She knows all about the property but does not know the *property itself*. The only way she can come to know the property itself is to become acquainted with that property by having the relevant experience. Becoming acquainted with the property in this way, then, doesn't mean acquiring any new information and so, if it is correct, the Knowledge Argument fails.

2.7 As mentioned, Alter (1998) has questioned the validity of drawing on the analogies of meeting a person or visiting a city in order to lend plausibility to the Acquaintance Hypothesis. Alter believes that it is a mistake to base the claim that acquaintance knowledge is *all* that Mary gains when she leaves the room and sees a colour on an analogy with what happens when one meets a person for the first time. This is because Conee fails to distinguish two kinds of cases of gaining acquaintance knowledge. Alter asks us to consider the following two examples of meeting the philosopher Rogers Albritton.

Albritton Case A. Everything I know about Albritton is based on reading his articles. Then I meet him, and this provides me with knowledge of what he is like. For example, I learn what he looks like and what his manner is;

and:

Albritton case B. Long before I meet Albritton, I see films of him (or films in which he is accurately portrayed by an actor). Seeing these films provides me with knowledge of what he is like – what he looks like, what his manner is etc. (1998, p.39)

Alter argues that it is not possible to model Mary's situation on either of these cases.

With regard to Case A, Alter claims that meeting Albritton would provide one with knowledge of certain facts about him, such as what he looks like and what his manner is, and learning what he is like would seem to consist in the acquiring of such facts. So if we were to model Mary's case on this scenario, it would support the idea that Mary must learn new facts on her release. Concerning case B, Alter is content to suppose that becoming acquainted with Albritton would indeed provide one with no new knowledge of facts about what he is like. Thus, were it to be applied to Mary's situation, we could conclude that learning what it is like to see a colour consists in gaining acquaintance knowledge only. But Alter says that in Case B we would know what Albritton is like *before* we become acquainted with him, which Mary cannot do with regard to colour, so Mary's situation cannot be aligned with Case B either. Alter concludes:

Thus, in certain cases gaining acquaintance provides one with knowledge of what that with which one becomes acquainted is like, and in other cases it does not. The cases in which it does are the cases like Albritton Case A – those in which one learns facts about the thing in question. Therefore, *contra* Conee, the analogy between the Mary case and *relevant* cases of becoming acquainted with a person or a city – namely, cases in which gaining acquaintance knowledge goes along with learning what the person or city is like – seems, if anything, to *support* the claim that she learns facts when released. (1998, p.40)

2.8 I believe there are two points that Conee makes, which he may wish to point out to Alter in answer to these objections. The first is with regard to Alter's Case A and the notion that Mary learns a new fact, such as 'what red things look like', when she first has the relevant experience. On this subject, Conee has the following to say:

Learning what red things look like is identical to learning how red things look, and this is identical to learning the look of red things. When we have reached this last formulation of the content of the learning, any appearance of factual content is gone. Clearly this does not say that what is learned is some fact to the effect that something or other is so. It says, concerning a certain look, that what is learned is *it*. A look is not a fact. This learning seems to be unproblematically classified as a relation of a person to a phenomenal quality, just as the acquaintance approach would have it. (1994, p.204)

Whether or not the above formulation can be accurately applied to Alter's Case A and learning what Albritton looks like the first time we meet him, it does appear to hold with regard to colour experience. This may seem at first to support Alter's complaint that Conee bases his theory on illegitimate analogies, but what I actually believe it does is show that Alter is mistaken to take Conee's use of such analogies too literally. I believe Conee uses the analogies of coming to know a person or a city more as an explanatory tool rather than an argument for the definite existence of acquaintance knowledge with regards to Mary's situation.

Conee also addresses the objection that by seemingly having complete physical knowledge about chromatic phenomenal qualities, Mary whilst still confined is *already* acquainted with them (Alter's Case B), so could already know what the experience is like. Conee here points out that the acquaintance hypothesis he is proposing denies that by knowing all the physical facts about phenomenal qualities one is acquainted with those qualities, as knowing all such facts does not imply *experiencing* them. He admits that this seems to be a special requirement for acquaintance with phenomenal properties and attempts to explain why someone who is thoroughly familiar with a particular place, say Cambridge, and who therefore 'knows Cambridge', can be said to be acquainted with the city, yet Mary's familiarity with the physical facts about chromatic vision is not sufficient for to know the associated phenomenal properties by acquaintance.

Conee's answer to this problem is to claim that acquaintance knowledge requires one to know the object of this knowledge directly; one cannot just be familiar with it, or know about it. When it comes to phenomenal qualities it is not sufficient for someone to know all such facts about the quality as we are capable of a more direct sort of awareness of that quality than can be gained by a conceptual representation.

In fact, however it is with the two other forms of knowledge [factual and ability], it seems particularly clear that knowing something by acquaintance requires a person to be familiar with the known entity in the most direct way that is possible for a person to be aware of that thing.

[...] The reason for this appears to be that whereas a person's awareness of such a [phenomenal] quality in knowing a fact about the quality can be mediated by a conceptual representation of the quality, we seem to be capable of a more direct sort of awareness of any such quality. When the quality is a property of someone's experience, the person need not use any such representation to be aware of the property. Perhaps awareness is experiential pure and unmediated; perhaps awareness of an experienced quality is mediated by some particularly transparent sensory form of representation. What matters for the present account is that experiencing a [phenomenal] quality is the most direct way to apprehend the quality. That much seems beyond reasonable doubt. (1994, p.207-8)

Despite evident confusion in Conee's position, I believe Alter's objection ultimately fails as Conee is not saying that knowing a particular phenomenal quality (of a physical property) by acquaintance is *exactly* analogous to knowing a person or a city, but rather it is a direct contact that can only be gained through experience.

Conee does however give himself a real problem which he fails to deal with, and it is to this that I now turn.

2.9 Conee cites Herbert Feigl (1958) as arguing that merely experiencing does not constitute knowledge of any sort, and claims that this is surely right if we take it to mean that a person's merely having some property or living through some condition is not sufficient to know the property or the condition⁹. Firstly, Conee points out that the crucial cases for the Knowledge Argument are relevantly dissimilar to this one, as they involve only knowledge of experiences. Then he says the following:

It is plausible that having experiences is sufficient for knowing those experiences. It is most plausible to hold that this is almost sufficient. The 'almost' is called for because qualities that are quickly and inattentively experienced may not be thereby known. Momentary peripheral awareness of some new shade of colour is not sufficient really to know that shade. The one thing more that is required in order to know an experienced quality is to *notice* the quality as it is being experienced. (1994, p.203-4: my italics)

It is the introduction of this idea of 'noticing' that is his undoing.

Conee does not believe that this 'noticing' clause implies any need for factual knowledge to be gained if we are to have knowledge by acquaintance. It is true that most of the time factual knowledge, as well as ability knowledge, will accompany such an experience, but it is not necessary in order for us to know what an experience is like. The problem is Conee does not give us any reason to accept this claim that we will not gain any new factual knowledge.

The addition of the stipulation that one must notice the experience as it is undergone protects Conee from accusations that qualities that are quickly and inattentively experienced may not be known, but what exactly does it mean to 'notice' an experience?

On this point Conee is not at all clear. He uses phrases such as ‘mere attentiveness will do’ and concludes that noticing an experience while it is undergone is all that is required in order to know what such an experience is like; but is such noticing essential? And if it is, does it really not result in additional factual knowledge?

Some may wish to try and argue that despite an initial inattentiveness to a particular experience it may be possible to remember such an experience when prompted at a later date. If one of Mary’s captors were to slip a piece of red paper under her door when she is gazing in that direction, yet so completely deep in thought concerning a particular problem that she fails to notice the experience, can we say with confidence that she would not gain the Lewisian abilities to recognise, remember and imagine and thus recall, when prompted at a later date, the experience of red that she did in fact have but failed to notice at the time? Were this to happen we would struggle to maintain that Mary does not know what an experience of red is like.

Of course the effectiveness of such a response would depend on what exactly is required to notice an experience. Conee may wish to claim (as indeed he does in the quote above when discussing the lack of need for conceptual representation) that it in no way requires any form of conscious introspection, and thus one could theoretically unconsciously notice an experience. But the point remains that Conee needs to clarify what exactly he means¹⁰. Any conscious introspection would seem to upgrade acquaintance knowledge to propositional knowledge of the kind ‘*this* is the experience I am having now’, which would of course cause problems for Conee’s theory. Perhaps the noticing is the noticing of the *colour*, rather than noticing the *experience* of the colour, ruling out any need for a form of higher order introspection which could lead to Mary getting new information. But then what exactly does such noticing entail? Such questions clearly need to be addressed¹¹.

More will be said on this notion of introspection in the next chapter when I consider the merits of the New Knowledge/Old Fact (NK/OF) view, but its surfacing here hints that the addition of this ‘noticing’ clause makes Conee’s acquaintance hypothesis suddenly look very similar to the idea that what Mary learns is actually an old fact under a new mode of presentation. This can be seen even clearer when he deals with another particular objection.

Conee concedes that when Mary experiences red for the first time, and thus becomes acquainted with phenomenal redness; she presumably becomes able to think about that property in a new way. Thus it may be objected that Mary can have certain accurate thoughts which are new, and new accurate thoughts imply new correct information about phenomenal redness – which in turn implies Jackson’s conclusion that there is correct information beyond that which Mary has access to within her black and white environment.

Conee seems to feel this objection can be dealt with fairly easily and does so with the following statement:

To have a new way to introduce the topic of a thought is to have a new means of referring to that topic. We have a new means of referring to a topic if we have a new symbol for that topic. Plainly the same thought can be newly symbolized. So the conceded representational difference by itself does not show that Mary has a new thought. (1994, p.209)

This explanation, then, does seem to point towards the use of different modes of representation called on by the NK/OF theories which, as we shall see in the next chapter, carries its own problems.

I feel there is a lot that is right about Conee's approach, but it stumbles as a response to Jackson's argument due to gaps in his explanation and its slide into the NK/OF view. In actual fact there is no real need for him to have aligned himself with the NK/OF view at all, as his initial hypothesis rests, by his own description, not on contents of representational thought but on simply noticing an experience whilst it is undergone. I shall come back to the Acquaintance Hypothesis in the conclusion and demonstrate its wider appeal, but for the moment it is Jackson's argument that we are trying to refute and the view Conee is presenting does not manage it.

2.10 So where have we got to? In this chapter I have shown that the Ability Hypothesis fails to solve the problem for physicalism as it requires too much of Mary in order for her to know what an experience is like. It seeks unsuccessfully to identify knowing what an experience is like with the possession of certain abilities. Next we have considered the Acquaintance Hypothesis, which tries to simplify matters and claims that to know what an experience is like is to have contact with such an experience in the most direct way possible – i.e. by having it, and noticing it as it is undergone – but to my mind has too many holes in it to successfully knock down Jackson's argument. The main difficulty surrounds exactly what is required of Mary to notice her experience as it is undergone. Any conscious introspection would seem undoubtedly to upgrade acquaintance knowledge to propositional knowledge, which then needs to be explained away by invoking the idea of different modes of representation, which brings its own separate problems, as we shall see in the next chapter. Is it possible to notice an experience without any need for some kind of representation, higher-order or otherwise? It may be so, but either way Conee needs to be clear about what he himself means.

Neither approach considered in this chapter manages to show that Mary can indeed come to know what an experience of colour is like without gaining any new propositional knowledge. So, if she does gain new *factual* knowledge, can this be reconciled with the truths of physicalism? This is the question to which I now turn.

3. Getting off at Q3: New knowledge of old facts.

3.1 If the new knowledge Mary seems to gain is neither ability knowledge nor acquaintance knowledge, we seem to have only one sort left: propositional knowledge. But if Mary gains new propositional knowledge after escaping from her achromatic prison, then Jackson would seem to have won and all physicalists should give up their position and make no attempt to disembark the Knowledge Argument train. What if,

however, this increase in propositional knowledge wasn't actually a *real* increase, but rather an *illusory* one? Hit the brakes driver, I'm getting off.

It's true that many physicalist philosophers find it hard to deny that Mary does gain new knowledge when she is released and sees a colour for the first time, and this knowledge is actually factual/propositional in nature. But rather than throw physicalism out the window, many are drawn to the New Knowledge/Old Fact view¹², which seems to give them an escape route.

This view draws on the resources of philosophy of language and also relies on a representational view of phenomenal consciousness¹³. The basic idea is that phenomenal character is a *physical property* of experiences, and to gain knowledge of what it is like to have an experience of a particular phenomenal character requires the acquisition of *phenomenal concepts* of phenomenal character. This can be explained in physical terms and phenomenal concepts can only be gained by experience. So Mary gains knowledge about phenomenal characters under phenomenal concepts but the facts that make these new items of knowledge true are physical facts that Mary knew before her release under another conceptualisation (a physical one). The important upshot of all this is that physical concepts and phenomenal concepts are thus *cognitively* (and conceptually) *independent* and can only be connected a posteriori, despite picking out the same (physical) fact.

Many philosophers take phenomenal concepts to be conceptually irreducible in the sense that they neither a priori imply, nor are implied by, physical-functional concepts. Controversy arises as antiphysicalists often take this further to imply that therefore phenomenal qualities are themselves irreducible and so are not physical functional properties. The point of the NK/OF view, I believe, is to show that the former need not imply the latter.

3.2 As well as providing a positive account of Mary's new knowledge that is consistent with physicalism, the NK/OF view has in its favour the fact that it gives an error-theory explanation of why the Knowledge Argument seems so intuitively appealing. As Nida-Rumelin points out:

In general, if a philosopher *A* claims that the argument of philosopher *B* does not go through, it is a point in favour of his view if he can provide an error theory, that is if he can explain why the argument may *appear* correct in the first place. The New Knowledge/Old fact View can claim to have an error theory with respect to the knowledge argument. Given the cognitive independence of physical and phenomenal concepts of blueness it *appears as if* we could imagine a situation where everything Mary knew before release were fulfilled but not what she came to know after release (and this can be taken to imply that she *does* come to know new facts). But, according to the NK/OF View this is an illusion. There is no such possible situation. What Mary learns after release is made true by a physical fact that she already knew before her release. (2002, p.12)

There are a number of different versions of the NK/OF view, but where the basic idea falls down is in its reliance on two distinct modes of presentation of a particular fact. The

approach, and the analogies often used to support it, appears to necessitate that at least one of the modes of presentation is contingent as they are unable to be connected a priori. It seems clear this is not the way phenomenal concepts work; to gain a phenomenal concept is not to re-learn something old in a new way, but to learn something new.

In this section I am going to consider one fairly standard approach from Michael Tye (1986, 2000), which – in addition to having other problems – does fall into this trap, and a less standard (and rather more sophisticated) approach from Brian Loar (1990), which apparently does not. This done, I shall not try to refute Loar’s approach directly but rather suggest that such arguments have rather missed the point of what the discussion should be about. In fact, I shall claim that what the NK/OF view really does is play into the hands of the dualists by focussing heavily on an unnecessary separation of the physical and the phenomenal.

3.3 As well as taking the NK/OF approach, Tye’s view seems to have the added bonus of accounting for the problems I identified in the last chapter with both the Ability Hypothesis and the Acquaintance Hypothesis. You will remember that the Ability Hypothesis failed when it became clear that it is both possible to have knowledge of what a particular experience is like without having any of the Lewisian abilities (to remember, imagine, recognise), and possible to possess the abilities, with regard to a certain experience, without knowing what it is like. As Tye puts it: “Cut the pie any way you like, then, the ability hypothesis is false” (2000, p.154).

In an initial move to fix the Ability Hypothesis Tye adds the ability to apply an indexical concept to the phenomenal character of an experience via introspection to Lewis’ list. He then rightly acknowledges that the addition of this ability doesn’t save the Ability Hypothesis as Lewis presents it (according to which to know what an experience is like just *is* the possession of these abilities), because one can have such an ability without knowing what a certain experience is like. To see this, consider a case where Mary is distracted enough, when shown a red rose for the first time, to stop herself introspecting the experience and applying any phenomenal concept; thus not knowing what it is like.

Now, as I discussed in the last chapter when considering the Acquaintance Hypothesis, it would seem that were Mary to not be distracted and attend to her experience in the way Tye proposes, she would indeed gain new propositional knowledge. This Tye accepts, and thus proposes the following adaptation to the Ability Hypothesis:

In the case described in the section above in which Mary is distracted, Mary has knowledge of how to do something. She knows how to mentally point to the phenomenal character of her experience in introspection. But, being distracted, she doesn’t exercise her know-how. Were she to do so, she would turn her know-how into knowledge-that. Intuitively, she would come to know that *that* is the phenomenal character of her experience. And in so doing, she would come to know what it is like to have an experience of that sort. So, the introspective knowing-that is sufficient for knowing what it is like. Such knowing-that is not necessary, however. One need not be paying attention to one’s current experiences to know what it is like to experience red. Intuitively, in such a case, it is necessary and sufficient to have abilities of the sort Lewis describes. It seems, then, that knowing

what it is like is best captured by a disjunction of introspective knowing-that and knowing-how along the following lines:

S knows what it is like to undergo experience E = df Either S is now undergoing E, and S has knowledge-that with respect to the phenomenal character of E obtained via current introspection, or S has the Lewis abilities with respect to E. (2000, p.155)

In Tye's theory, then, we seem to have answer to the problems we encountered in the previous chapter with both the Ability Hypothesis and Acquaintance Hypotheses. While gaining certain abilities remains an important part of the normal way of things, one no longer needs to identify it with knowing what an experience is like. Also, we can use Tye's requirement of applying an indexical concept to the phenomenal character of an experience via introspection to fill the gaps left by the Acquaintance Hypothesis of just what it means to 'notice an experience as it is undergone'. Conee may well wish to protest that this is not at all what he meant, but in the absence of any other explanation it seems to me a reasonable suggestion, and one that fits in nicely with the overall picture.

The problem Tye now has, of course, is that if Mary is not in a position to gain the necessary Lewisian abilities, perhaps due to problems with her visual imagination, she seems to have to apply an indexical concept via introspection and thus acquire new propositional knowledge in order to know what an experience is like. This would seem to entail that she *cannot* therefore know all the facts whilst confined within her room; thus Tye would have to concede physicalism is false. Tye of course does no such thing, and this is where he brings the NK/OF view into play.

Tye argues that whilst in the room Mary lacks the phenomenal concept *red*, which does not mean that Mary attaches no meaning to the term 'red' – she will be able to use the term correctly in many different cases – only that the concept she exercises is *non-phenomenal*. Knowing what it is like to experience red is necessary for possession of the phenomenal concept *red*. Tye does not see Mary's lack of such phenomenal concepts whilst in her room as grounds for the claim that she therefore cannot know all there is to know about colour vision prior to her release. Tye denies that where a difference between the old and the new concepts obtains, a difference in the world between the properties these concepts stand for or express must also obtain – conceptual differences need not be mirrored in worldly differences.

When Mary leaves her room, has and introspects her first experience of red she will have the thought *that she is having an experience with this phenomenal character*. Tye admits that Mary cannot think this thought truly whilst in the room as the concept *this* refers to the phenomenal character associated with an experience of red; thus it seems again that when she leaves the room and thinks this thought she is making a genuine discovery. Tye deals with this problem by appealing to the role modes of presentation play in the individuation of thought-contents.

In this sense, what I think, when I think that Cicero was an orator, is not what I think when I think that Tully was an orator. This is precisely why it is possible to discover that Cicero is Tully. The thought that Cicero was an orator differs from the thought that Tully was an orator not at the level of truth-conditions – the same singular proposition is partly constitutive of the content of both – but at the level

of concepts and modes of presentation. The one thought exercises the concept *Cicero*; the other the concept *Tully*. The concepts have the same reference; but because they present the referent in different ways, the two thoughts can play different roles in rationalising explanation. (2000, p.157)

So, basically what we have are two different modes of presentation which can only be connected a posteriori. The physical concept and the phenomenal concept both pick out the same (physical) property, but within her room Mary does not have access to the phenomenal concept and thus only *seems* to gain new knowledge upon her release. In actual fact she knew all the facts before her release, she just effectively re-learns an old fact in a new way.

3.4 Before getting on to some specific objections regarding the different modes of presentation, there seem to be some obvious confusion in Tye's position. He claims that in order to know what an experience is like one must either be applying a phenomenal concept to the experience as it is undergone or gain the requisite Lewisian abilities. This would suggest that were Mary to be distracted in the way described above, and thus not attend to, say, an experience of red in the way Tye proposes, she may gain the Lewisian abilities but not have acquired the phenomenal concept *red*. In such a case can she really be said to know what it is like to experience red¹⁴?

It seems quite possible that were Mary to acquire the Lewisian abilities in this way she may not know she has them until she has cause to exercise them, and surely until she does so she is no more the wiser about what it is like to experience red than before she left the room. Moreover, what is it that will happen when she does finally *visualise* red, say? How is it she will come to know what such an experience is like? Surely, Tye would have to say it will be when she introspects such imaginings and applies the necessary phenomenal concept.

Maybe it is indeed possible to gain the Lewisian abilities without introspecting the initial experience, but I do not see how one can be said to know what the experience is like until one has exercised those abilities and done it *consciously* which, for Tye, involves the application of an indexical concept.

It seems that Tye's approach does not need to include the Lewisian abilities, and he'd do better to focus on the need to gain a phenomenal concept via introspection. This, of course, then requires him to explain away the seeming increase in propositional knowledge, which he does by claiming that rather than learn new facts, a person would come to know old facts in a new way – via a different, phenomenal, mode of presentation.

Torin Alter (1998) objects to the NK/OF approach in the same way he did to Conee's Acquaintance Hypothesis, by highlighting that the kind of analogies appealed to to support this view, on closer inspection, fail to do so. He presents one such analogy as follows:

Jane knows that Muhammad Ali is fifty-four years old, but she does not know that he used to go by the name 'Cassius Clay', a name she has not previously heard. She then reads in *Sports Illustrated* the statement 'Cassius Clay is fifty-four years

old'. She takes herself to have learned a new fact about a boxer's age. However, the fact that Clay is fifty-four years old and the fact that Ali is fifty-four years old are one and the same. It is the fact that a certain man has a certain property; the man is Ali, a.k.a. Clay, and the property is *being fifty-four years old*. Jane knew that fact *before reading Sports Illustrated*. (1998, p.41)

If we view the Mary case in a similar fashion, what is happening when she has certain experiences is that she is adding 'phenomenal guises', as Alter calls them, to her representational repertoire; guises unavailable to her whilst she is in the room despite knowing all the physical facts. When she is released she comes to apprehend the old facts under these new phenomenal guises, in the same way that Jane apprehends the old fact that Ali is fifty-four under the new 'Clay-guise'. Why Alter believes this analogy fails is as follows.

Basically, in the Jane example, we can say that the reason she knows the fact expressed by the sentence 'Cassius Clay is 54 years old' before she ever hears the name Clay is that so far as knowing that fact is concerned it makes no difference what representation is employed so long as it represents the right man, Ali – there is nothing special about the name Clay. But Alter does not believe this is the same in Mary's case and the phenomenal guises she acquires *do* seem to be special in just this respect. Alter is saying that the phenomenal guises are not *incidental* to knowing facts about what colour experiences are like in the way that pseudonyms are incidental to knowing facts about what someone's age is. To claim that Mary can know such facts before acquiring the phenomenal guises is like claiming that before Jane reads the sports magazine she is in a position to know that Ali was once called Clay. But in order to do so Jane would have to have 'Clay' in her vocabulary, and in the same way, for Mary to know the facts at issue she would have to have the relevant phenomenal guises already in her representational repertoire. As Alter describes:

One of the facts relevant to the Mary case is, in Old-Fact/New-Guise terminology, that the experience of seeing red has the phenomenal guise G. The Old/Fact/New-Guise analysis depends on the claim that it is possible to know *that* fact even if, as in the case of the pre-release Mary, G is not in one's representational repertoire. That claim is doubtful. G figures into the fact that seeing red has G so centrally that it seems implausible that one could know it without having G at one's cognitive disposal. Claiming otherwise is like claiming, implausibly, that one can know the fact that a certain boxer was once called 'Clay', even if one has neither heard nor read the name 'Clay'. (1998, p.41)

So the experience of red has the phenomenal guise G, and Ali can be known under the alternate guise 'Clay' – is claiming that Mary can know the experience of red has G without ever experiencing red, the same as claiming that Jane can know Ali is Clay, without ever hearing the name 'Clay'? Does Tye's appeal to such an analogy require him to accept the seemingly absurd conclusion that Jane could know Ali is Clay?

Alter's reason for thinking this is that phenomenal concepts are not incidental to knowing facts about what a colour experience is like in the way a certain name can be incidental to knowing facts about a certain person. As he says above, it is implausible to think that

Mary can know an experience of red has the phenomenal guise G before having an experience of red, but is this really the case?

Tye (1986) confronts this type of objection with regards to his example of Jones, the blind neuroscientist who recovers his sight. Tye claims Blind Jones (like Mary) will know of the existence of the phenomenal quality of a particular experience. He will be aware that such an experience causes a phenomenal experience in others (such as his friend Smith) and that such a phenomenal quality has a designator etc, so Jones has all the facts that someone else who has had the experience does. Thus, whatever knowledge Jones gains after the operation to restore his sight it is not knowledge of any new facts.

One of the things Jones knows of *e* is that it has the phenomenal content which is typically caused in Smith by his viewing red objects. Jones is also aware of Smith's rigid name, '*R*', and he knows a description which fixes its referent, namely 'the phenomenal content which is typically caused in Smith by his viewing red objects'. Hence, Jones knows that *R* is the phenomenal content which...etc. Hence, Jones must surely know of *e* that it has *R*.

[...] factual knowledge of *e* and its phenomenal content is either knowledge of *e* that it has *R*, where '*R*' is a rigid designator for the phenomenal content, or it is knowledge of *e* that has the *F*, where 'the *F*' is a description for the content linking it to various causes or effects. Jones even before he gains his sight has both the former knowledge and the latter, whatever '*F*' signifies. Hence, there is no fact Smith knows about *e* and its phenomenal content of which Jones is ignorant. (1986, p.12-13)

This type of debate is difficult to adjudicate over. It is not completely obvious what exactly Alter's G is referring to. It does seem clear that Mary can know that an experience has a particular phenomenal character, but whether she can fully *understand* such a concept is questionable. In any case Alter's objection shows up why any use of a Cicero/Tully, Ali/Clay type of analogy is not relevant as phenomenal concepts clearly do not refer to the relevant properties in the way names like Ali and Clay do. This point is about to become even clearer when I consider Loar's approach.

3.5 Let us take a step back for a moment and consider the form of Jackson's argument, which Loar (1990) and Churchland (1985) have attacked on the grounds that it cannot be unrestrictedly valid, when presented as follows:

Since physical and phenomenal properties can be connected only a posteriori, physical properties must be distinct from phenomenal properties.

Loar provides a straightforward counterexample to Jackson's form:

- 1)Max knows the bottle in front of him contains CH₃CH₂OH
 - 2)Max does not know the bottle in front of him contains alcohol.
- Therefore
- 3)Alcohol CH₃CH₂OH¹⁵

Clearly the conclusion does not follow from the premises. However the antiphysicalist has a reply to this, which Loar anticipates. The antiphysicalist can concede that the Knowledge Argument thus interpreted is indeed not valid, and that property identities can be true even if not a priori, but only if one or more of the descriptions conceives of the reference via a contingent mode of presentation (as in the Cicero/Tully, Ali/Clay examples). They can then argue that this is not how phenomenal concepts work. Loar illustrates this by asking us to consider Kripke's famous argument concerning pain:

Kripke assumes that a phenomenal concept such as 'pain' cannot be a priori linked with a physical concept such as that of the stimulation of C-fibres. The case of Mary is a vivid way of making the same point. Kripke points out that property identities can be true even if not a priori, for example, 'heat = such and such molecular property'. It seems fair to represent the next step in his argument as follows. 'Heat' has a contingent higher-order mode of presentation that connotes the property 'feeling like this'. That is what accounts for the a posteriori status of the identity. But, as Kripke points out, this cannot be how 'pain' works: the phenomenal concept 'pain' does not pick out its referent via a contingent mode of presentation; it conceives pain directly [...] and essentially. Kripke concludes that pain is not identical with a physical property. (1990, p.224)

Tye's approach claims that Mary gains knowledge of an old fact under a new mode of presentation, but Kripke argues that this could only be the case if it were a *contingent* mode of presentation akin to the Ali/Clay example. Loar's problem is that he, like Alter, argues that the Mary case, unlike the Max case above or the Ali/Clay case, does *not* involve a contingent mode of presentation of something she has otherwise known all along. Mary, according to Loar, has a *direct* grasp of the property involved in the new information:

[...] there is no contingency in Mary's conception of the new phenomenal information that explains it as a novel take on old facts. She learns new facts simpliciter and not new conception of old facts.(1990, p.223)

So how then does Loar propose to explain away the seeming increase in propositional knowledge? Loar explains that both Kripke's and Jackson's arguments are based on the same implicit assumption we have just mentioned: that the only way to account for the a posteriori status of a true property identity is if one of the terms express a contingent mode of presentation. Or alternatively as:

(Semantic premise) A statement of property identity that links conceptually independent concepts is true only if at least one concept picks out the property it refers to by connoting a contingent property of that property. (1990, p.224)

So, two assumptions are at the foundation of Jackson's and Kripke's arguments: that phenomenal concepts and physical-functional concepts are conceptually independent; and the semantic premise. Loar's strategy is simple – he accepts the former assumption, but rejects the latter.

The antiphysicalist intuition that links concept-individuation and property-individuation (more closely than is in my view correct) is perhaps this. Phenomenal concepts and theoretical expressions of physical properties both conceive their references essentially. But if two concepts conceive a given property

essentially, neither mediated by contingent modes of presentation, one ought to be able to see a priori – at least after optimal reflection – that they pick out the same property. Such concepts' connections cannot be a posteriori; that they pick out the same property would have to be transparent.

But as against this, if a phenomenal concept can pick out a physical property directly or essentially, not via a contingent mode of presentation, and yet be *conceptually independent* of all physical-functional concepts, so that Mary's history is coherent, then Jackson's and Kripke's arguments are ineffectual. We could have two conceptually independent conceptions of a property, neither of which connote contingent modes of presentation, such that substituting one for the other in an opaquely interpreted epistemic context does not preserve truth. Even granting that our conception of phenomenal qualities is direct, physicalism would not entail that knowing the physical-functional facts implies knowing, on an opaque construal, the phenomenal facts; and so the failure of this implication would be quite compatible with physicalism. (1990, p.224-5)

That we should view phenomenal concepts in this way, according to Loar, is down to their being recognitional concepts that pick out certain internal properties; these are physical-functional properties of the brain. To have the phenomenal concept *red* is to be able to recognise red experiences while having them. The recognitional concept *red* refers *directly* to its referent, which is the physical property, and there is no other property (no property of that property) involved in the reference fixing. Loar is saying that the recognitional concept of phenomenal redness refers to the physical property phenomenal redness in virtue of being 'triggered' by that property.

Thus the knowledge argument seems to fail for the same reason in the case of Mary that it does in the case of Max, as both arguments require substitution in opaque contexts of terms that are conceptually independent. The difference between the two cases is that in the Max example the conceptual independence arises from a contingent mode of presentation, where as in the Mary example it arises from elsewhere.

3.6 In this chapter we have tried to get off the Knowledge Argument train a little further down the line than we did with the Ability Hypothesis and Acquaintance Hypothesis. We granted that Mary did indeed gain new knowledge and that that knowledge was of the propositional kind – only the facts she gained new knowledge of were either old ones she already knew or new phenomenal facts about physical properties in a way that is compatible with physicalism.

The tack taken by Michael Tye and others is shown to be mistaken in the analogies it employs to support its case, and would only work if the phenomenal concept conceived of the physical property via a contingent mode of presentation, which phenomenal concepts clearly do not do. Otherwise it would seem we should be able to connect the physical and the phenomenal concepts a priori, which we cannot do.

Loar then took up the baton and attempted to demonstrate that it is possible for physical concepts and phenomenal concepts to be conceptually independent, non-contingent and still not be able to be connected a priori (due to phenomenal concepts being recognitional concepts), thus showing that Mary can gain genuinely new propositional knowledge in a

way that is compatible with physicalism and renders Jackson's argument basically invalid. Whether he's successful or not seems difficult to say.

It certainly seems that the knowledge Mary gains is genuinely new, as opposed to merely rehashed under a different conceptualisation, as Tye is proposing. Van Gulick (2004) describes Loar's view as Mary learning new propositions, but the propositions are individuated in a fine-grained way. Were physicalism to leave out propositions of a coarse-grained kind, its claim to be a comprehensive account of reality might well seem in doubt, but if it only leaves out propositions of a fine-grained kind, perhaps it is ok. Either way we are likely to have reached an impasse, with supporters of the knowledge argument maintaining that Mary learns new propositions in a coarse-grained sense and those trying to refute it saying she learns new propositions in the fine-grained sense. Van Gulick says it is difficult to see how to resolve the conflict and I agree with him.

The main problem with the NK/OF view is not in fact found in the minutiae of semantic complexities, but rather the overall concept of the approach. An over reliance on philosophy of language, which Jackson's set up encourages, actually lends unnecessary support to the anti-physicalist cause. The focus on the conceptual independence of physical concepts and phenomenal concepts, and the discussion of phenomenal facts as well as physical facts, can surely only leave dualists rubbing their hands with glee, whilst suggesting the easy step up to the total independence of physical *properties* and phenomenal *properties* that they are arguing for.

Mary is said by Tye to gain, via introspection, phenomenal concepts of phenomenal character, which are representations of physical facts already known under a physical/theoretical conceptualisation. Loar claims the phenomenal concepts are a type of phenomenal fact, that conceive of a physical property directly or essentially in the same way as a physical-functional concept. As well as this unnecessary separation of the phenomenal and the physical, the equating of physical facts and phenomenal facts seems to sideline what the argument is supposed to be about – experience. It is not at all clear what the importance of *concepts* is in *experience* and a total focus on them seems move away from what's really important. Many other questions also arise as to what exactly a phenomenal *fact* would be, and whether they could really exist, but I'm getting onto the wider problems of the knowledge argument which I shall return to in the conclusion.

Now though I'm going to take a bit of a step back and consider a view from way down at the other end of the scale.

4. Jumping off at Q1: The Dennettian rebellion.

4.1 As mentioned earlier some of the most die hard physicalists try to get off the Knowledge Argument train very early on. The problem is that at this point the train has only just reached full speed with no scheduled stop in sight. Needless to say, anyone jumping off at this point is set to die a quick death¹⁶. Unsurprisingly this option has not proved very popular.

When faced with the question of whether Mary learns anything after leaving her black and white room, such philosophers simply shout ‘No!’, and assume no more needs to be said. Mary can indeed learn everything there is to know about colour vision from within a black and white room and this includes knowing what it is like to see colours. To many of us such a claim just seems so absurd as to not merit any attention at all, but in fairness to those who support so radical a view, they may or may not be completely mistaken about Mary’s epistemic situation, but they are right to point out that one should be wary of rejecting a view based solely on the fact it is counterintuitive.

As so often when it comes to contentious approaches, leading the march is philosopher Daniel C. Dennett. Whilst rejecting Jackson’s argument as just another *intuition pump*, Dennett (1991, 2006) argues that we have no good reason to believe that Mary cannot know what it’s like to experience red whilst inside her room and much reason to think she can, so Jackson’s argument poses no problems whatsoever. In this next section I shall show that in fact Dennett is perfectly correct to suggest that Mary could possibly know what it is like to experience a colour without ever stepping out of her room, but this fact does not in anyway defeat the knowledge argument, and certainly doesn’t prove Dennett’s point that experience does not teach us anything new. Dennett, like many philosophers, has missed the point of what the argument is all about and thus focussed on all the wrong bits.

4.2 The main thrust of Dennett’s objection seems to be, straightforwardly, that there is no more reason for accepting that Mary learns something new when she sees a colour for the first time than for accepting that she doesn’t; except perhaps a certain amount of traditional intuition which should be discarded. To illustrate this point he presents an alternate ending to the Mary example that he believes has as much claim to the truth as Jackson’s original.

And so, one day, Mary’s captors decided it was time for her to see colours. As a trick, they prepared a bright blue banana to present as her first colour experience ever. Mary took one look at it and said “Hey! You tried to trick me! Bananas are yellow, but this one is blue!” Her captors were dumbfounded. How did she do it? “Simple,” she replied. “You have to remember that I know *everything* – absolutely everything – that could ever be known about the physical causes and effects of colour vision. So of course before you bought the banana in, I had already written down in exquisite detail, exactly what physical impression a yellow object or a blue object (or a green object, etc.) would make on my nervous system. So I already knew exactly what *thoughts* I would have (because, after all, the ‘mere disposition’ to think about this or that is not one of your famous qualia, is it?). I was not in the slightest surprised by my experience of blue (what surprised me was that you would try such a second-rate trick on me). I realise it is *hard for you to imagine* that I could know so much about my reactive dispositions that the way blue affected me came as no surprise. Of course it’s hard for you to imagine. It’s hard for anyone to imagine the consequences of someone knowing absolutely everything physical about anything!” (1996, p.60)

Most of us will immediately appreciate how such a suggestion flies totally in the face of our natural intuitions about such cases. “If Mary has never seen a colour, how on earth could she possibly know the banana’s blue and not yellow?” I would expect that many materialists might allow that if Mary hooked herself up to some kind of brain scanner and

objectively monitored the results whilst having the experience of blue she may be able to empirically deduce that she is in fact having an experience of blue and not of yellow, but does this entail that she can know what an experience is like whilst never having had it? Such a suggestion seems absurd.

Dennett is claiming that Mary, using her vast knowledge, can work out what it would be like for her to see something blue, in advance of having such an experience, and arguing that we have no reason to think she could not. In the example above, Mary claims she has written down in 'exquisite detail' what impression an object of a particular colour would make on her nervous system. Many philosophers, in the face of this example, resort to restating the ineffability of qualia: commenting on the greater richness of phenomenal experience that inevitably transcends any description and so Mary *must surely* learn something, but need this be the case?

Lycan formulates the point thus:

Here's a way to see why some of us think Mary does learn something. What one knows when one knows w.i.l. [what it's like] to experience a blue sensation is ineffable; at least, it's very tough to put into (noncomparative) words. One resorts to the frustrated demonstrative: "It's like...*this*." The reason physically omniscient Mary doesn't know what it's like is that the ineffable and/or the ineliminably demonstrative can't be deduced, or even induced or abduced, from a body of impersonal scientific information. (As quoted by Dennett: 2006)

In the face of an argument like this, Dennett simply questions the traditional assumption that what one knows, when one knows what it's like to experience blue, is in fact ineffable. He mentions the ready effability of what it is like to experience a triangle and how an author can convey what it's like to see Paris by moonlight, and asks if we can really be sure that what it is like to experience colour can't be conveyed to one who has never done so in a few million words. What is it about colour experiences that makes such descriptions impossible, acknowledging the fact that Mary knows *everything* there is to know about colour that can be learned by physical science? Her exceptional abilities will surely include the ability to take in vast quantities of theoretical information, which would give limitless bounds to any description.

Dennett is clearly correct when he points out to Lycan – in reference to the latter's claim that an experience is 'very tough' to convey in words – that we are dealing with impossibility not difficulty, but that does not necessarily help his position. In order for Mary to really know what an experience of blue is like, the description would seem to have to include information over and above the language of physics, but this would then only support the antiphysicalist claims. The whole point of the knowledge argument is that talk of chemicals and neurons cannot convey the phenomenal character of experience, so some other kind of description is needed. If Dennett is claiming that Mary can write down a description in the language of physics that allows her to know what an experience is like then he is just begging the question in favour of physicalism.

It would seem that the only way Mary can recognise that the banana is blue without ever leaving the room would be to effectively by-pass the experience and gain some kind of

objective (non-phenomenal) access to her central nervous system and monitor changes in the same way she would through some kind of brain scanner. Whether or not such a feat is logically possible (in fairness who knows what abilities, *complete* knowledge would afford), all it would actually do is give Mary the ability to *recognise* a colour experience, which is of course not the same thing as knowing what it is like (another nail in the coffin of the Ability Hypothesis). When given the blue banana Mary may well be able to say ‘hang on chaps, this banana is blue’, but she still wouldn’t have known what an experience of blue is like until they actually gave it to her. If Dennett thinks otherwise he clearly has misunderstood what the phrase means¹⁷.

4.3 Actually, there may well be another way Mary could recognise the experience of blue, and indeed know what it is like without ever leaving her black and white room, and that is to experience it whilst she is in there (yes, I know what you’re thinking, but bear with me). Dennett uses the example of RoboMary¹⁸, to show why we should accept that Mary could easily put all her scientific knowledge of colour to use and figure out exactly what it would be like to experience colour. He presents it as follows:

1. RoboMary is a standard Mark 19 robot, except that she was brought on line without color vision; her video cameras are black and white, but everything else in her hardware is equipped for color vision, which is standard in the Mark 19.

2. While waiting for a pair of color cameras to replace her black-and-white cameras, RoboMary learns everything she can about the color vision of Mark19s. She even brings colored objects into her prison cell along with normally color-sighted Mark 19s and compares their responses—internal and external—to hers.

3. She learns all about the million-shade color-coding system that is shared by all Mark19s.

4. Using her vast knowledge, she writes some code that enables her to colorize the input from her black and white cameras (à la Ted Turner’s cable network) according to voluminous data she gathers about what colors things in the world are, and how Mark19s normally encode these. So now when she looks with her black-and-white cameras at a ripe banana, she can first see it in black and white, as pale gray, and then imagine it as yellow (or any other color) by just engaging her colorizing prosthesis, which can swiftly look up the standard ripe-banana-color-number-profile and digitally insert it in each frame in all the right pixels. After a while, she decides to leave the prosthesis turned on all the time, automatically imagining the colors of things as they come into focus in her black and white camera eyes.

5. She wonders if the ersatz coloring scheme she’s installed in herself is high fidelity. So during her research and development phase, she checks the numbers in her registers (the registers that transiently store the information about the colors of the things in front of her cameras) with the numbers in the same registers of other Mark 19s looking at the same objects with their color camera eyes, and makes adjustments when necessary, gradually building up a good version of normal Mark 19 color vision.

6. The big day arrives. When she finally gets her color cameras installed, and disables her colorizing software, and opens her eyes, she notices . . . nothing. In fact, she has to check to make sure she has the color cameras installed. She has learned nothing. She already knew exactly what it would be like for her to see colors just the way other Mark 19s do. (2006)

Anyone who shares Dennett's position that a human being is basically some kind of biological machine should accept this analogy. They should accept that it is logically possible that Mary, with her immense knowledge, could, in principle, do for herself what RoboMary does and thus come to know what it is like to see colours whilst in her room. But does such an admission defeat the Knowledge Argument? No. Dennett is just adding further confusion over what it means to have an experience in the first place.

Whether or not Mary could give herself colour experiences whilst still confined within her black and white environment has nothing to do with whether or not she learns something new upon actually having such an experience. As soon as RoboMary colourises the input from her black and white cameras, she has a colour experience and thus learns what it's like to see colours; gaining knowledge she did not have access to before she had done so. Once she has learned about the colourising system of the other Mark 19s, she could be said to know all the physical facts about colour experience, but until she puts such knowledge into practise in herself she will not have a colour experience. The physical knowledge itself will not tell her what it is like to experience red or green or blue, which, of course, is exactly Jackson's point.

The Knowledge Argument does not hinge on Mary being unable to have a colour experience within her room, just that until she has such an experience she cannot know what it's like, despite having access to all the physical/theoretical knowledge there is. RoboMary's appreciation of the exact code that makes her fellow robots see colours is not the same as seeing a colour herself and to believe this is to deny that a colour experience has an experiential side which cannot be conveyed accurately in talk of light waves and computer circuits. That is to say, it is to totally eliminate phenomenal conscious experience from our existence. Some do indeed choose to do this, but I consider such a belief to be simply nonsensical (though I don't intend to argue that here).

4.4 In case there was still any doubt as to Dennett's misunderstanding of the whole point of the knowledge argument, here is another example he presents in support of his point.

Swamp Mary: Just as standard Mary is about to be released from prison, still virginal re colors ... a bolt of lightning rearranges her brain, putting it by Cosmic Coincidence into exactly the brain state she was just about to go into *after* first seeing a red rose. (She is left otherwise unharmed of course; this is a thought experiment.) So when, a few seconds later, she is released, and sees for the first time, a colored thing (that red rose), she says just what she would say on seeing her *second* or *nth* red rose. "Oh yeah, right, a red rose. Been there, done that."
(2006)

From this Dennett draws the conclusion that you can come to know what an experience is like without ever having had the experience. Once again it is clear how Dennett is mistaken. Once Mary's brain has been rearranged, she has an experience of red: simple as that. And thus she now knows what such an experience is like. Having her brain rearranged into the state that it would enter into were she to see a red object, is exactly the same, as far as experience of that colour goes, as actually seeing the object. That just seems to be to be an obvious physicalist assumption.

4.5 To conclude, then, Jackson's argument is that Mary cannot know what it is like to experience red until she actually *has* such an experience. No amount of physical information can change this. It is an important point (also made by Churchland (1985)) that we can really have no idea of what capabilities an entire knowledge of neuroscience would bring and Dennett's entertaining examples show that it may be logically possible for Mary to recognise red and actually have an experience of red, without ever actually coming into contact with a red object. Well OK that may well be true – but it does not refute the Knowledge Argument.

Blue Banana Mary may learn to observe her own nervous system in some weirdly objective way, but does not know what an experience of blue is like until the blue banana is put in front of her. Both Mary and RoboMary may be able work out exactly what is happening in their human/robot counterparts when they have an experience of red, they may be able to play around with their own internal equipment to give themselves such an experience despite only having access to black and white inputs; but until they instigate such changes in their own brains (circuits?) the information alone does not teach them what it is like to experience red. To claim otherwise is to deny our entire experiential reality. And that really is just silly.

Conclusion: Stopping the train before it starts.

5.1 Let's recap. At the beginning of this essay I laid out the knowledge argument as follows:

- 1) Mary has all the *physical* information about colour vision before her release
- 2) However, there is some information about human colour vision that she does not have before her release (namely – she does not know what it is like to experience colour)

Therefore

- 3) Not all information is physical information.

Given the wide acceptance of premise one, the focus of the responses has been either to show that Mary doesn't learn anything new when she first experiences colour, or to accept that she does and then explain away the increase in knowledge in a way that is compatible with physicalism. Has either been successfully done by any view we have considered? In the introduction I posed a number of questions which helped to divide the responses into different camps. Let's now go over them briefly one by one.

Q1. Does Mary actually learn anything new or gain any knowledge when she leaves her room and first has an experience of colour?

Most of the philosophers discussing the knowledge argument happily concede that Mary does indeed learn something when she has her first experience of colour – but not all. As we have seen, Dennett tries to convince us that by having her first experience of colour,

super-informed Mary learns absolutely nothing at all. The way he tries to do this is by explaining how Mary could either recognise her first colour experience by some kind of third person observation of her nervous system, or give herself an experience of colour whilst still in the room. Whether or not either of these scenarios is logically possible, as we have seen, is totally beside the point.

Mary being able to recognise a particular experience of colour by monitoring chemical changes in her body by some non-phenomenal means is not in itself going to tell her what such an experience *is like* (in the way Nagel's term must surely be intended). As for Mary being able to put herself in the correct physical (brain) state to have an experience of colour in the absence of any external stimuli, what does that show? External stimuli or not, before she puts herself into such a state she does not know what it is like to experience colour. Once she has done so, she gains the knowledge of what it is (phenomenally) like, knowledge she didn't have before – thus Dennett's arguments are ineffectual.

On to Q2 then.

Q2. If Mary does gain new knowledge, is it *factual/propositional* in nature?

The theories which answer 'no' to this question rely on Mary gaining a different kind of knowledge to propositional knowledge; some kind of non-factual ability knowledge or acquaintance knowledge. Lewis leads the charge on ability knowledge but we saw him fall when it was shown you could know what an experience is like without possessing any of the necessary abilities and vice-versa. Conee's acquaintance hypothesis tries to build on Lewis' failure and simplify what is enquired in order to know what an experience is like. His view however, in addition to looking suspiciously like a version of the NK/OF approach, really doesn't explain anything at all. He says that we can know what an experience is like simply by noticing it as it is undergone, but doesn't explain what he means by 'notice'.

So can we explain away Mary's apparent increase in knowledge by appealing to a different kind of non-factual knowledge? On the strength of the views considered here, we would have to say, no.

On to Q3 then.

Q3. If it seems Mary does gain propositional knowledge, does she gain knowledge of *new* propositions or just come to know *old* propositions in a new way?

From here on it all gets a bit complicated. The assumption is that the new knowledge Mary gains is indeed of a *fact*. The problem is Mary is supposed to know all the facts already. The answer: what she actually gains is knowledge of an old fact she already knew, but in a new way (under a different mode of presentation). I have looked at two approaches for how this could work.

Tye gives the fairly standard Cicero/Tully type argument that has the physical property as the sole referent of both a physical concept and a phenomenal concept; same physical fact known in two different ways. These views are rejected by philosophers of language on the grounds that the only way two conceptually independent concepts could conceive of the same property and not be able to be connected a priori (which of course phenomenal and physical concepts cannot), is if one of the concepts conceives of the property via a contingent mode of presentation, which is not the way phenomenal concepts work. Phenomenal concepts and physical concepts both refer to the physical property directly and essentially and thus should be able to be connected a priori.

Loar argues that this can be explained by accepting that phenomenal concepts are recognitional concepts and as such can refer directly to the physical property without needing to be able to be connected to the physical concept a priori. Therefore, Mary can indeed gain some form of new propositional knowledge but the knowledge is of the same physical property designated by the physical concept so is no threat to physicalism.

Is the knowledge argument now refuted? It is difficult to say. It seems quite open to philosophers of language just to deny that phenomenal concepts are recognitional concepts or claim that the new knowledge is knowledge of a phenomenal property rather than a physical property. Indeed the NK/OF view does seem to support the existence of separate phenomenal properties by discussing the acquisition of phenomenal concepts, not of physical properties but of phenomenal character. This phenomenal character is then in some way eliminated by discussing phenomenal concepts referring directly to the physical property itself. Invoking phenomenal concepts as being important in discussions of experience merely draws attention to the divide between the physical and the phenomenal in a way that is clearly unhelpful to the physicalist cause.

5.2 So, the Knowledge Argument train is still rumbling along and at the moment we seem powerless to stop it. Does this mean we should finally accept defeat and abandon physicalism for good? Not at all, because once we realise what physicalism *really* means, we can stop the train from moving off in the first place.

Jackson's argument relies on the premise that physicalism entails that Mary has access to *all* the physical information (and that is obviously the only information there is) about colour experience whilst confined within her black and white environment. All we need do to rid ourselves of the Knowledge Argument is to deny that this is what physicalism means. Simple.

When we say that Mary can have access to all the information from within her black and white room, what exactly do we mean? We mean that she has access to all the *theoretical* information – everything described in the language of *physics*. Is there any real reason for us to believe that such theoretical information is all there is to know about the nature of the physical world? Must we accept that a belief in physicalism commits one to the view that physical reality in its entirety can in principle be captured in the language of physics?

This seems to be the position of most physicalist philosophers as demonstrated by the responses to Jackson's argument. The Mary case suggests that we can refute physicalism by demonstrating that an omniscient colour scientist can know all the physical facts about colour experience without knowing what it is like to experience colour. Faced with this challenge the physicalist philosophers have come flying out the blocks to demonstrate, not that it isn't true that someone can know all there is to know about colour experience without actually ever seeing a colour, but that its truth doesn't entail the falsity of physicalism. They happily accept that physicalism entails Mary can gain complete knowledge of the physical nature of colour experience from within her black and white environment. They also accept that this complete knowledge still leaves her in the dark about what such experiences are like. And yet they still believe physicalism can be saved.

Well, this may or may not be the case (though from the positions dealt with in this paper it doesn't look hopeful), but the truth is the Knowledge Argument isn't arguing against physicalism at all – as we shall now see.

5.3 Galen Strawson (2003, 2006) has argued that this view held by the vast majority of physicalists, is actually not physicalism at all, but rather should be known as *physicSalism* and dismissed. As he explains:

It follows that real physicalism can have nothing to do with *physicSalism*, the view – the faith – that the nature or essence of all concrete reality can in principle be fully captured in the terms of *physics*. Real physicalism cannot have anything to do with *physicSalism* unless it is supposed – obviously falsely – that the terms of *physics* can fully capture the nature or essence of experience. (2006, p.1)

In his discussions of the so named 'mind-body problem', or the 'problem of experience', Strawson – unlike most physicalists – starts from the truth of the experiential (traditionally described, misleadingly, as the *mental*) rather than that of the non-experiential (or *physical*). This is because he acknowledges that it is the experiential whose existence is 'more certain than the existence of anything else'. Real physicalism (unlike that espoused by eliminativists like Dennett), he claims, must accept first and foremost that experiential phenomena exist and, second, that they are wholly physical phenomena. If we are ever going to progress in the study of consciousness/experience; if we ever hope to bridge the so called 'explanatory gap' – solve the 'hard problem' – we must begin from the experiential and work backwards to the physical.

Strawson explains how physicalist philosophers who accept the doctrine of *physicSalism* are guilty of an inflated belief in how much we know and understand of the nature of the physical. The reason philosophers, then, find discussion about the experiential so difficult is because it doesn't fit in with everything we *know* about the physical, but, as Strawson says of the many philosophers who have fallen into this trap:

They have already made a large and fatal mistake. This is because we have no good reason to think that we know anything about the physical that gives us any reason to find any problem in the idea that experiential phenomena are physical phenomena. (2006, p.1-2)

The truth is we only *really* know very little, so the fact that what we *do* know is portrayed in a way that seems to create a conflict between our physical knowledge of the world and our experience of it, should not concern us. It certainly should not lead us to accept dualism or, even more hysterically, eliminate experience altogether.

This essay has shown how an over reliance on physicalism can actually only promote dualism, or require us to eliminate the experiential. If we take theoretical physics as one thing and experience as another, we push experience out of the physical picture. We treat mental and physical as two totally separate things, as if we are talking about cows and animals as opposed categories (as Strawson puts it). The experiential is as much a part of our physical reality as rocks, atoms and everything else, and we need to treat it as such.

Never are these deficiencies in the ongoing debate more obvious than in discussions of the Knowledge Argument. The fact that it focuses on *knowledge* creates instant confusion as philosophers assume this means only theoretical knowledge. Suddenly the language of physics becomes a stand in for the 'physical' and the experiential gets lost somewhere amongst phenomenal facts/concepts/properties – totally separated and often eliminated. Before they know it the philosophers involved have only emphasised the supposed mental-physical divide and perpetuated the myth of dualism in the very act of trying to refute it. With this happening it seems almost impossible that we can ever escape from the old dualist intuitions.

Strawson explains that it is indeed true that a real physicalist must accept that the mental/experiential is physical, it is just neurons firing, but accepting this does not mean that everything that is going on can be described in the terms of physics.

When I say that the mental, and in particular the Experiential, is physical, and endorse the view that "experience is really just neurons firing," I mean something completely different from what some materialists have apparently meant by saying such things. I don't mean that all aspects of what is going on, in the case of conscious experience, can be described by current physics, or some nonrevolutionary extension of it. Such a view amounts to radical "eliminativism" with respect to consciousness, and is mad. My claim is different. It is that the Experiential (considered just as such) – the feature of reality we have to do with when we consider experiences specifically and solely in respect of the Experiential character they have for those who have them as they have them – that "just is" physical. No one who disagrees with this is a remotely realistic materialist. (2003, p.51-52)

Why should we accept Strawson's view of what physicalism should be? Strawson ultimately argues the case for panpsychism¹⁹, but it is beyond the remit of the current essay to decide whether one should endorse what may seem a fairly extreme position. It suffices for me just to demonstrate that there is no good reason for us to accept that physicalism should entail physicalism, and thus no reason to accept that a belief in physicalism commits one to the claim that Mary can know everything there is to know about colour experience whilst in her black and white environment. Once this claim is rejected we have no need to try and explain away her seeming increase in knowledge.

Strawson does seem in strong company with regards to our lack of knowledge of the physical. He quotes Russell as saying: “Physics is mathematical, not because we know so much about the physical world [non-mental, non-experiential] but because we know so little” (2006, p.5). Also Eddington: “Our knowledge of the nature of the objects treated in physics consists solely of readings of pointers [on instrument dials] and other indicators [...] science has nothing to say about the intrinsic nature of the atom” (2006, p.5)

Whether or not Strawson’s ultimates (the fundamental parts of the universe) are wholly and utterly non-experiential or wholly experiential, the real point for me is that there is just no reason to believe the complete nature of the physical world can be captured in the language of physics. Whether some form of expression will be developed in the distant future which can capture the essence of the physical world is another issue, and not one I have time to discuss here. Suffice to say that any views that maintain that ultimates are totally non-experiential appear to slide into dualism or eliminativism. If in order to maintain that everything in the world is physical and no non-physical stuff exists we have the choice to either give up the common view that ultimates are non-experiential, or the very existence of our experiential reality – it certainly seems more sensible to do the former.

5.4 If we accept this new definition of physicalism, and I believe we should, Jackson’s argument has no power against it.

So what does Mary find out when she leaves her room for the first time and encounters a red rose or green grass? Well, she has an experience of red, or green and finds out what it is like to have such experiences. From within her room she knew that when she encountered a red rose light would be reflected into her eye at a certain wavelength instigating a process of chemical reactions in her brain and central nervous system which ultimately gives her an experience of red²⁰. She knew before she ever had the experience that it consisted of such-and-such a combination of neurons firing. However, the neurons in *her* brain had never yet fired in such a combination. The light of such-and-such a wavelength had never entered *her* eye. Knowing the theoretical physics of experience does not give you the knowledge of what that experience is like – only having the experience will do that.

From this I seem to be heading in the direction of some form of Acquaintance Hypothesis, and indeed this is also Strawson’s preferred approach:

[...] we have direct acquaintance with – know – fundamental features of the mental nature of (physical) reality just in having experience in the way we do, in a way that has no parallel in the case of any non-mental features of (physical) reality. We do not have to stand back from experiences and take them as objects of knowledge by means of some further mental operation, in order for there to be acquaintance and knowing of this sort: the having is the knowing. (2003, p.54)

So what we are acquainted with through our experience is reality *as it is in itself*. Whether or not this can be captured in the terms of physics or anything else is irrelevant. The

reason Mary does not know what it is like to experience red whilst confined in her room is simply because she is yet to do so: the having is the knowing.

Granted, the hypothesis presented by Conee (1994) that we considered earlier stumbled as a response to Jackson's argument, but that is no longer a reason for dismissing it as a possible solution to the problem of experience as we have despatched Jackson's argument by other means. In any case, Conee's argument failed when he attempted to account for the idea that Mary could quickly or inattentively experience some particular colour and therefore not actually come to know what it is like – and this takes us into a representational debate that is ongoing with no clear end point.

Exactly how to characterise the kind of *knowledge* Mary gains does indeed pose a problem, but such difficulties need not make us question the truth of *real* physicalism. Perhaps we can just label it *experiential* knowledge – knowledge of experience that is contained in *having* the experience. Discussions of phenomenal consciousness²¹ often make the phenomenal nature of our existence out to be so mysterious when it's obvious to see why it exists. I expect much has been written on the evolutionary benefits of the short-hand information system that phenomenal experiences such as pain provide.

What is important for the purposes of this essay is that Jackson's argument can be refuted, and this can be done by denying that physicalism entails Mary can know everything there is to know about colour experience without ever actually having an experience of colour. If Jackson's argument does succeed, it only does so against physicism, not physicalism – and nobody should believe in that.

5.5 As a final point, one thing Jackson's argument certainly does succeed in doing is highlight the sheer amount of confusion that is currently suffocating debates in philosophy of mind. Progress is being hindered largely due to the terminological chaos, exemplified in this essay by the difficulties with tired phrases such as Nagel's *know what it's like*. Though most philosophers would claim to understand what Nagel meant by this term, this in no way means everyone is using it in the same way. Similar problems arise elsewhere.

The Knowledge Argument is particularly affected by such problems with issues over exactly what is meant by a *fact* – physical or phenomenal – and how this relates to the idea of *information*. Add to this the 'what it's like' problem and issues surrounding representational theories of consciousness and it can sometimes feel that attempting to refute the argument is akin to trying to solve an algebra equation without any numerical values whatsoever!

Progress in this area can only be made once the terminological issues are addressed and arguments such as Jackson's, which serve only to mislead the debate, are pushed to one side.

Notes

¹ Though there has been some discussion of weak and strong versions of Jackson's argument, it is generally taken that 'information' here refers to propositional knowledge or physical *facts*.

² There are of course notable exceptions which I shall come back to, as it is my own preferred route to a conclusive refutation of Jackson's argument.

³ Similar approach used by Van Gulick (2004).

⁴ One particular approach not covered is (now ex-dualist) Jackson's recent attempt to refute his own argument. Having concluded that physicalism must in fact be true and embraced the representational theory of perceptual experience, Jackson blames his prior acceptance of the Knowledge argument on the illusory belief in the property of phenomenal redness. An illusion now dispelled by the representational theory, which he, interestingly combines with the Ability Hypothesis:

Those who resist accounts in terms of ability acquisition tend to say things like "Mary acquires a new piece of propositional knowledge, namely, that seeing red is like *this*", but for the representationalist there is nothing suitable to be the referent of the demonstrative.

We have ended up agreeing with Laurence Nemirow and David Lewis on what happens to Mary on her release. But, for the life of me, I cannot see how we could have known they were right without going via representationalism. (Jackson 2003, p. 439)

Jackson's objection to his tricky thought experiment is not particularly compelling. It is completely reliant on the truth of the representational theory and it is not completely clear why his representationalism leads him to the Ability Hypothesis. I have omitted it here in order to focus on the main strands of objections that have been put forward.

⁵ A similar approach is put forward by Lawrence Nemirow (1990).

⁶ This does not of course mean Lewis believes that actually having an experience is the only *possible* way of coming to know what it is like to have it (it could, say, be produced by magic or extremely advanced neurosurgery), but it would seem that in this world it is the only practical way. This distinction will actually become more important later on when discussing Dennett's position.

⁷ Lewis defines parapsychology as: "the science of all nonphysical things, properties, causal processes, laws of nature, and so forth that may be required to explain the things we do." (1988, p.93)

⁸ It may well be argued that this isn't actually true at all and we can indeed know what a person or a city is like before we meet them/visit it. Conee himself acknowledges a distinction between *experiencing* a city and a colour, as we discover later on in our

analysis of his position. As well as highlighting inconsistencies in Conee's view and problems with the analogies he uses, this issue is an early sign of the confusion that results due to misappropriation of Nagel's phrase.

⁹ The example Conee gives to illustrate this is of a man who wins the lottery, but falls into a coma before he knows about it and all the money is eaten up by hospital bills. Thus, when he awakes he does not, and never did, know what it is like to be wealthy.

¹⁰ This again reflects the terminological problems surrounding exactly what is required in order for a subject to know what an experience is like.

¹¹ Such debates are raging currently amongst representationalist theorists, with no obvious solution in sight. Thus I fancy a simple answer to Conee's problem of just what it is to 'notice' an experience is unlikely to be forthcoming. Its emergence here also demonstrates the similarity between Conee's views and those that come under the New Knowledge/Old Fact position, which will be discussed in the following chapter.

¹² Hereafter referred to as the NK/OF view.

¹³ The representational view of phenomenal consciousness is by no means widely accepted, though I personally feel it has a lot going for it. Either way I do not intend to argue for or against it here.

¹⁴ Once again we find confusion lurking within Nagel's phrase.

¹⁵ Similar examples are used by Churchland (1985), including showing temperature is not identical to mean kinetic energy of the molecules, and water does not equal H₂O.

¹⁶ Ok, so I've pushed the metaphor a little far here, but you get the point!

¹⁷ More confusion over Nagel's phrase – never!

¹⁸ Dennett accepts that some may find the comparison with robots to be invalid as they could not have colour experiences, but then his view is that a belief in materialism entails that it should be possible in principle to build a robot-brain that can do just what our brains can do. Thus anyone who writes off such a possibility is assuming the falsity of materialism and thus begging the question in favour of dualism. Dennett however seems to be begging the question in favour of physicalism, emphasising another problem that there may well be no neutral starting point in these debates.

¹⁹ Which Strawson takes to be: The view that the existence of every real concrete thing involves experiential being. He believes that ultimates (the fundamental parts of the universe) cannot be wholly non-experiential.

²⁰ Obviously my knowledge of such a process is somewhat inferior to Mary's.

²¹ Yet another disputed term, but one which I take to refer to our experience.

References

- Alter, T. 1998. "A Limited Defense of the Knowledge Argument", *Philosophical Studies*, 90, 35-56
- Alter, T. 2001. "Know-How, Ability, and the Ability Hypothesis", *Theoria*, 67, 229-39.
- Broad, C. D. 1925. *The Mind and its Place in Nature*, London: Routledge and Kegan Paul.
- Churchland, P. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States", *Journal of Philosophy*, 82, 8-28.
- Churchland, P. 1989. "Knowing Qualia: A Reply to Jackson." In his *A Neurocomputational Perspective*, 67-76. Cambridge: MIT Press. Reprinted (along with a 1997 postscript) Ludlow, et. al. (2004), p.163-178
- Conee, E. 1994. "Phenomenal Knowledge", *Australasian Journal of Philosophy*, 72, 136-50. Reprinted in Ludlow, et. al. (2004), p.197-216
- Dennett, D. 1991. "Epiphenomenal Qualia?" In his *Consciousness Explained*, Boston: Little Brown and Company, pp. 398-406. Reprinted in Ludlow, et. al. (2004), p.59-68
- Dennett, D. 2006. "What RoboMary Knows." Forthcoming in T. Alter and S. Walter (Eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. New York: Oxford University Press. Available: <http://ase.tufts.edu/cogstud/papers/RoboMaryfinal.htm>
- Feigl, H. 1958. *The "Mental" and the "Physical"*. Minneapolis: University of Minnesota Press.
- Horgan, T. 1984. "Jackson on Physical Information and Qualia", *Philosophical Quarterly*, 34, 147-52. Reprinted in Ludlow, et. al. (2004), p.301-308
- Jackson, F. 1982. "Epiphenomenal Qualia", *Philosophical Quarterly*, 32, 127-36.
- Jackson, F. 1986. "What Mary Didn't Know", *Journal of Philosophy*, 83, 291-5. Reprinted in Ludlow, et. al. (2004), p.51-56
- Jackson, F. 1995. "Postscript", in *Contemporary Materialism*, ed. by Paul K. Moser and J. D. Trout, New York: Routledge, 184-9. Reprinted in Ludlow, et. al. (2004), p.409-416
- Jackson, F. 1998. "Postscript on Qualia." In his *Mind, Method, and Conditionals: Selected Essays: 76-79*. London: Routledge. Reprinted in Ludlow, et. al. (2004), p.417-420
- Jackson, F. 2003. "Mind and Illusion", in Ludlow, et. al. (2004), p.421-442
- Lewis, D. 1983. "Postscript to 'Mad Pain and Martian Pain.'" In his *Philosophical Papers*, vol. 1. New York: Oxford University Press, 130-32.
- Lewis, D. 1988. "What Experience Teaches", *Proceedings of Russellian Society (University of Sydney)*, Reprinted in Ludlow, et. al. (2004), p.77-104
- Loar, B. 1990. "Phenomenal States (Revised Version)", in *The Nature of Consciousness: Philosophical Debates*, N. Block, O. Flanagan, G. Guzeldere, eds. Reprinted in Ludlow, et. al. (2004), p.219-240

-
- Ludlow, P., Stoljar, D., and Nagasawa, Y. eds. 2004. *There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. Cambridge: MIT Press.
- Nagel, T. 1974. "What Is it Like to Be a Bat?", *Philosophical Review*, 83, 435-50.
- Nemirow, L. 1990. "Physicalism and the Cognitive Role of Acquaintance", in *Mind and Cognition: A Reader*, ed. by William G. Lycan, Oxford: Blackwell, 490-9.
- Nida-Rumelin, M. 2002, "Qualia: The Knowledge Argument", *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/qualia-knowledge/>
- Stoljar, D., and Nagasawa, Y. 2004. Introduction to Ludlow, et al. (2004), p.1-36
- Strawson, G. 2003. "Real materialism", in *Chomsky and his Critics*, ed. L. Antony and N. Hornstein (Oxford: Blackwell) 49-88.
- Strawson, G. 2006, forthcoming. "Realistic monism: why physicalism entails panpsychism", *Journal of Consciousness Studies*, 13: 1-24
- Tye, M. 1986. "The Subjective Qualities of Experience", *Mind*, 95, 1-17
- Tye, M. 2000. "Knowing What It Is Like: The Ability Hypothesis and the Knowledge Argument." In his *Consciousness, Colour, and Content*: 3-20. Cambridge: MIT Press. Reprinted in Ludlow, et. al. (2004), p.143-160
- Van Gulick, R. 2004. "So Many Ways of Saying No to Mary". In Ludlow, et. al. (2004), p.365-406