Lewis on Consciousness and Qualia

Should a Materialist Believe in Qualia?
The paper with this title is an attempt to show how much Lewis can accept of what the qualia theorist says without giving up materialism. It’s a useful exposition of Lewis’s philosophical methodology, particularly in its focus on moderation. But I think he concedes too much here, and the materialist need not even risk the epithet ‘eliminativist’. Or, at least, she need not risk this for the reasons that Lewis worries about here.

Lewis thinks that there are two distinctive features of the folk psychological role of qualia. First, experiences give us new abilities. As Lewis says, one part of the concept of qualia is that they are “properties of experiences apt for causing abilities to recognize and to imagine experiences of the same type.” (r327) As we’ll see below, Lewis thinks that this feature of the folk psychological role is crucial to explaining (away) the intuitions behind Frank Jackson’s ‘knowledge argument’.

Qualia can be identified when we have them. And identified in a ‘particularly strong sense’. Here is one version of the strong thesis Lewis wants to defend.

If I have an experience with quale Q, I know that I am having an experience with quale Q. (r327)

On the next page, Lewis elaborates this bit of folk theory.

The knowledge I gain by having an experience with quale Q enables me to know what Q is – identifies Q – in this sense: any possibility not ruled out by the content of my knowledge is one in which it is Q, and not any other property instead, that is the quale of my experience. Equivalently, when I have an experience with quale Q, the knowledge I thereby gain reveals the essence of Q: a property of Q such that, necessarily, Q has it and nothing else does. If, for instance, Q is essentially the physical property of being an event of C-firing, and if I identify the qualia of my experience in the appropriate ‘demanding and literal’ sense, I come to know that what is going on in me is an event of C-firing. (r328)

Lewis calls this the Identification Thesis. Although he doesn’t believe it, he thinks it is a bit of folk psychology, and acknowledges it as a cost of materialism that we have to give it up. I think this is wrong twice over. First, for broadly Williamsonian reasons, folk theory is just as committed to the denial of the Identification Thesis as its affirmation. Second, there is little reason to attribute the Identification Thesis, as opposed to one of its near cousins, to the folk.
Williamson has argued at length that none of our states are *luminous*. That is, there is no state such that when we are in it, we know that we are in it. Williamson’s argument is often presented somewhat technically, but in fact all of the premises that he uses are parts of folk theory. Here is the argument.

1. Knowledge requires safety; that is, that if there is a saliently similar situation to the actual one in which \(\neg p\), the agent does not know that \(p\).
2. Borderline cases are everywhere; that is, for any state \(S\) there is a possible agent \(a\) and situation where \(a\) is in \(S\) but there is a similar situation in which she is not in \(S\).
3. So, for any state \(S\), it is not the case that necessarily all agents who are in \(S\) know that they are in \(S\).

In both 2 and 3 the modal claims are made with respect to a quite restricted modality. The borderline cases Williamson describes obtain in worlds very similar to the actual world. So even if the Identification Thesis is only meant to hold in worlds similar to our own, Williamson’s argument would tell against it.

It might be suggested at this point that folk psychology endorses the Identification Thesis despite the presence of this argument. In that case folk psychology is inconsistent. And the materialist is well within her rights to insist that the best way of making it consistent before using it to define terms is to drop the Identification Thesis.

A better objection is to worry that folk theory isn’t really committed to premise 1 in Williamson’s argument. In “Luminous Margins” I argue (following Mark Sainsbury) that the right safety constraint is a little different to this one. As I note there however, premise 1 is still true in a number of cases. In particular, if \(p\) is the proposition that I am in state \(S\), and state \(S\) is a state that I come to know I’m in by anything like observation, then premises 1 and 2 are true for those substitution instances of \(p\) and \(S\). By ‘anything like observation’ I mean to include any kind of perceptual or introspective knowledge. So we can restate Williamson’s argument as follows.

1. Observational knowledge requires safety; that is, if the agent comes to believe that \(p\) on the basis of observational evidence, then if there is a saliently similar situation to the actual one in which \(\neg p\), the agent does not know that \(p\).
2. Borderline cases are everywhere; that is, for any state \(S\) there is a possible agent \(a\) and situation where \(a\) is in \(S\) but there is a similar situation in which she is not in \(S\).
3. So, for any state \(S\) such that the agent comes to know she is in \(S\) by observational means, it is not the case that necessarily all agents who are in \(S\) know that they are in \(S\).

The restriction to observational knowledge is relevant because it is actually quite hard to tell whether our knowledge of our own qualia is best understood as coming about via introspection or
coming about because these states are self-presenting. If the states are self-presenting, then knowledge that we are in them is not, in the relevant sense, observational. On this model, we don’t ‘look inside’ and see that we are in pain. Rather, the pain just is (or at least includes) an awareness of being in pain. The thought then is that the safety condition comes from thinking about cases of observational knowledge (Williamson’s examples are things like judging the heights of trees of the size of football crowds) and it isn’t appropriate in cases where there is no distinction between the state itself and our judgment that we are in the state.

So does this resurrect the Identification Thesis? I think not. As I said, it’s really hard to see whether we should be using some kind of introspective model of experiential self-awareness, or something like a self-presentational model. The fact that it is hard to see how to even start modelling the problem suggests that folk psychology is silent on the matter, and hence that folk psychology doesn’t commit to the self-presentational model. But if the Identification Thesis were part of folk psychology, then it would be so committed, because the restated Williamsonian argument rules out an observational model of self-awareness. So even when we take account of the possible problems with Williamson’s argument, we see a reason to deny that folk psychology is committed to the Identification Thesis.

The observations in the previous paragraph about folk psychology’s reticence are directly borrowed from the discussion of the Language of Thought hypothesis in “Reduction of Mind”. As we’ll see presently, there is an important methodological lesson of that discussion. We shouldn’t think that folk psychology, or for that matter any folk theory, is committed one way or the other on hard philosophical questions. To the extent that our arguments imply a commitment of folk theory, that is a cost of the arguments. (Compare also Lewis’s discussions of the argument from temporary intrinsics, where he regards it as a cost of his own view that it implies the folk are committed to perdurance rather than endurance.)

Bracketing for now these Williamsonian concerns, let’s set out the argument that the Identification Thesis is incompatible with materialism. As far as I can tell, it goes this way.

1. If the Identification Thesis is true, then if the quale of my experience is Q, I can know that it is Q.
2. If materialism is true, then for some of my experiences, the quale is the property of being an event of C-firing.
3. I am never in a position to know that one of my experiences is an event of C-firing.
4. So, if the Identification Thesis is true, materialism is false.

Premise 3 perhaps attributes more ignorance to me than is really warranted, but we can assume for the sake of this argument that I don’t know much about neurology, so I’m not in a position to know about C-firings. So we’ll grant premise 3. But why should we grant premise 2?
Here is one argument that Lewis might be using. If materialism is true, then all properties are physical properties. The physical property that is best suited to being the quale of a pain experience is the property of being a C-firing. So the quale of a pain experience is the property of being a C-firing. (As Lewis says, being a pain experience might just amount to being an experience that is a C-firing, so the quale might add very little.) And that seems to get us to 2.

But the problem with this argument is that the first premise is false. It doesn’t follow from materialism that all properties are physical properties. Some properties are functional properties that happen to be realised physically. (Perhaps this is terminological; maybe these functional properties are best called physical properties as well. If you’d prefer to talk that way then what I say here should count as a reason that the quale is the property of being a C-firing.) And it seems natural to say that qualia are functional properties. At least this seems just as natural as saying that they are physical properties. Intuitively I could have had an experience with the very same qualia without having any of the same physical properties. Perhaps this intuition should be discarded, it does seem dangerously close to the anti-materialist that intuition I could have the same physical properties without the same qualia. But I think for now best to accept it, and easily save the Identification Thesis.

Lewis might well have had a more subtle argument in mind. For instance, in “Reduction of Mind” he stresses the argument that mental properties can be causally efficacious, but functional properties (in the sense I’m using the term) are not causally efficacious, so mental properties are not functional properties. As argued in previous notes, the right response to this argument is that once we look closely at the relations between properties and events, we can keep the intuition that mental properties are causes, and that they are functional properties, without having in the bad sense disjunctive causes. The trick is to note that the theory of events already rules out disjunctive causes, so we can let our mental properties be as disjunctive as we like. If they get disjunctive, then the event of having some mental property won’t essentially be a having of that property, and so won’t be a disjunctive cause.

I think this is a perfectly sufficient response to the anti-materialist argument from the Identification Thesis. But it’s worth pausing for a little over some potential ambiguities in the statement of the Identification Thesis. Remember that in “Reduction of Mind”, Lewis allows that even if ‘pain’ denotes a physical property, he’d allow that ‘being in pain’ denotes a functional property, i.e. the property of having the property that plays the pain role. This presumably generalises. Just like we can talk about an experience being Q, we can talk about experiences having the property of being Q. These will come apart in the following way.

Q is the property of an experience that explains why we can re-identify experiences of that type. For pains, that property is being C-firings. So an experience is Q iff it is a C-firing. An experience has the property of being Q if it has the property, whatever it is, that explains why we can re-identify experiences of that type. It is a priori (but contingent) that an experience of mine has the property of being Q iff it is Q. This is related to the fact that it is a priori (but contingent)
that an object has the property that plays the heat role iff it is hot. Given that, if we can tell that our experiences have (essentially) the property of being Q, we’ll be able to tell that they are Q whenever they are, using our direct knowledge that they have the property of being Q, and our ability to infer that whatever has the property of being Q is Q.

It might be thought that this will take us back to the original problem. After all, we just said that we can know that our experiences are Q. But the original point was that we can’t know this. But thinking again through how we got to this point will show us what was wrong with the anti-materialist argument.

Start with a case removed from issues about materialism. A child puts his hand on a stove. He comes to know what qualia he has, but he also comes to know that the stove is hot. (I’m not claiming here any kind of necessary connection between knowledge and proximity to hot things. I’m just saying that as a matter of fact, some of the time that children put their hands on hot stoves, they come to know that the stove is hot.) Following what Lewis says above, we might try paraphrasing this by saying that every possibility consistent with what the child knows is a possibility in which the stove is hot.

Now let’s think about heat for a moment. Lewis wants to say that the stove is hot iff it has the property, in fact rapid molecular motion, that plays the heat role. The child is causally effected by the stove in virtue of it having this property, and so is in a position to know that the stove has that property, whatever it is, that plays the heat role. From this, he can correctly infer that the stove is hot. So we seem to have worked ourselves into the same kind of problem as the friend of qualia was pressing on the materialist, but without the qualia. (Remember the focus here is on the stove’s heat, not on the qualia this produces in the child.) Since this problem must have a solution, the matching problem for qualia must have a solution.

I’d prefer to solve both problems using two-dimensionalist resources. This is not quite going to be a Lewisian solution, since Lewis’s preferred use of the 2D apparatus involved interpreting so-called diagonal propositions metalinguistically, while I’ll follow Chalmers in interpreting them epistemically. But it’s in the same spirit.

The possibilities that we rule out when we come to learn something are not possible worlds, but epistemic possibilities. Epistemic possibilities differ from possible worlds in many ways. The most important is that they do not respect essential properties. If it’s essential to potassium that its atomic number is 17, then in any possible world, its atomic number is 17. But nevertheless, there are (or would be) epistemic possibilities in which its atomic number is 18. An epistemic possibility is a plurality of sentences that is (a) closed under logical implication and (b) contains all truths of which we can be a priori certain, and (c) does not contain some sentence. (Some people use ‘known’ rather than ‘certain’ in (b), but this leads to odd results if certain kinds of rationalist epistemologies are accepted.) Some of the sentences in an epistemic possibility are indexical. For instance, an epistemic possibility may contain the sentence “I am Queen Elizabeth the First”. That’s one of the many possibilities I know I don’t inhabit.
Now, when the child learns puts his hand on the stove, he gets to rule out all the possibilities in which the stove lacks the property that plays the heat role from being live possibilities for him. Simultaneously, the gets to rule out all the possibilities in which the stove is not hot. But this is no cognitive achievement, since the epistemic possibilities in which the stove is hot just are the epistemic possibilities in which it has the property that plays the heat role, since it is a priori (but contingent) that an object is hot iff it has the property that plays the heat role.

And that, perhaps belatedly, is I think the best materialist response to the argument from the Identification Thesis. Perhaps folk psychology says that when we have an experience with a certain quale, we know exactly which quale that is. But perhaps that isn’t right, perhaps we just know which quale the experience has the property of having. We can infer which quale the experience has, but in doing so we’ll be ruling out epistemic possibilities, not ruling out the *metaphysical* possibilities in which our experience is not, as it might be, a C-firing. I seriously doubt that folk psychology takes sides in a debate between the theorist who says that when we have an experience with a quale, we know what it is, and a theorist who says that in that case we know that the experience has the property of having the quale. Folk theory tends not to be quite so fine-grained. But the latter theorist says nothing that a materialist need reject. So a materialist is in a position to accept everything on which the folk theory of qualia is committed.

**Know How and Ability**

Lewis’s response to the knowledge argument is what is commonly called the Ability Hypothesis. When Mary comes out of the black and white room, she does not gain any propositional knowledge, but she does gain some abilities. In particular, she gains the ability to imagine having a certain kind of experience. Let F be a property that characterises the kind of experiences an agent would have were she to eat a Vegemite sandwich. If the agent has never tasted Vegemite, then she won’t be able to imagine having an F experience. Indeed, she might not be able to imagine this even if she knows what experiences are F. But those of us fortunate enough to have tasted Vegemite can do this with ease. We can also imagine other related experiences, such as the experience of having Vegemite on toast, or eating Vegemite mixed with sugar.

Lewis mentions some other abilities that Mary gains on leaving the black and white room, but these seem less essential to the response. He says that we gain the ability to recognise when we are later having the same experience. Well, in one trivial sense of course that’s right. If I don’t taste Vegemite, then I can’t later recognise having the same experience, because I don’t have the same experience. Lewis must mean something stronger, but it isn’t quite clear what this is. I think the idea is that in virtue of the experiences I get when I first taste Vegemite, I gain the ability to recognise taste V, so when I next taste Vegemite, I will recognise that it is V, and hence that it is the same experience as I previously had. But even if I hadn’t tasted Vegemite the first time, I would still recognise when I taste it the second time that its taste is V. So it isn’t clear what extra is gained when I taste Vegemite the first time is any distinctive kind of ability to
re-identify, since the relevant recognitional capacities would have been triggered by the second taste.

Perhaps what is crucial here is the fact that it is easier to remember that we have had experiences with certain qualia than it is to remember other properties of our experiences. That is true, and it is an ability that is associated with qualitative properties of our experiences. But we can’t infer from this that there is any ability that we gain when we have a new kind of experience. If F is a qualitative property, then when I have an experience that is F, I’ll be able to remember that one of my experiences is F, and I wasn’t able to do that before. But that isn’t anything distinctive of qualia. (Indeed, saying that it is is Lewis’s third way to miss the point.) What seems to be the case is that I have a general ability to remember which qualitative properties my experiences have instantiated at some time or other, and when I have an experience with a new qualitative property, I trigger that ability to remember that they’ve had this qualitative property.

An analogy might help here. Suzy is an extremely fast runner. But there are no baseball teams in her area, so in one sense she lacks the ability to steal bases, because to steal bases one must be playing baseball, and she doesn’t have the ability (given her situation and resources) to play baseball. When she joins a baseball team, she now has the ability to steal bases. But it is misleading to say she acquired the ability to steal bases. Rather, a change to the external circumstances gave her a new way of applying her old ability of running very fast, namely applying this speed to stealing bases. Similarly, when I have an experience with qualitative property F, I have the ability to remember that my experience is F, but I think this is best thought of as a new application of my old ability to remember the qualitative properties of my experiences. This is not to reject the Ability Hypothesis, just to insist that the abilities in question should be taken to be primarily imaginative.

We should also distinguish the Ability Hypothesis from a close cousin with which Lewis frequently confuses it. The close cousin is the Know How Hypothesis. Lewis explicitly says that what Mary gains is know how, not know that, but the official story is that she gains neither of these, she gains an ability. Now there is a difference between ability and know how. Rickey Henderson knows how to steal bases. That’s why he’s the greatest base stealer in history, and (I imagine) he’ll be a great base running coach. But Rickey is old, and his legs won’t carry him as fast as they used to. So Rickey lacks the ability to steal bases. If he had that ability, he might still be on a major league baseball team, but sadly he is not.

So in principle there is a difference between ability and know how, and it is possible that Mary might gain an ability when she sees red without even gaining any know how. If Rickey’s legs were returned to their former power, either by surgery or chemistry, then he would again have the ability to steal bases. But neither the surgery nor the chemistry would increase Rickey’s knowledge, not even his know how. I’m inclined to think that the Ability Hypothesis is somewhat more plausible than the Know How Hypothesis, and from now on I’ll be keeping the two firmly separate.
One reason for keeping the Ability Hypothesis distinct from the Know How hypothesis is that it is arguable that know how is a kind of know that. (Jason Stanley and Timothy Williamson have argued just this.) But it isn’t really arguable that an ability just is a piece of know that. So we should preserve the focus on abilities, not knowledge how.

New Mode of Presentation Response

It is worth distinguishing the Ability Hypothesis from its closest rival materialist response. This is the New Mode of Presentation response, which says that Mary ‘learns’ some old information about the physical world, but learns it under a new guise.

Sometimes when we learn old information under a new guise, it can seem like we are learning something new. For instance, if we previously knew that Hesperus is visible from Rome, and we then ‘learn’ that Phosphorus is visible from Rome, then it might seem that we’ve learned something new. But in fact, if the planet itself is a constituent of the proposition that we learned, then this is something we already knew. We don’t have to posit some fact about Phosphorus that isn’t also a fact about Hesperus in order to explain why this seems like learning.

Could this kind of picture be applicable to Mary as well? Well, not obviously. It’s true in principle that she could have two representations of the same underlying facts about the world, her mind, and the relation between them. And it’s even true that having an experience with a qualitative feel could produce a different kind of mental representation than could, in practice, be produced without the experience. But there’s a crucial difference between Mary and the Hesperus/Phosphorus case that seems to defeat this response.

If we know that Hesperus is Phosphorus, then when we acquire the mental state that represents that Phosphorus (as such) is visible from Rome, we won’t think we’ve acquired new information. The impression that gaining a new mental representation is a learning experience only lasts as long as we don’t know that the new representation represents something we haven’t already represented some other way. Now Mary is meant to know all there is to know about colour perception. That means that she knows, inter alia, how phenomenal properties are represented in the brain, and what they are representing. That means that she knows, inter alia, that the new representation she gets when she sees red for the first time is a representation of an old fact. So she shouldn’t think that she is learning.

Still, there is something right about this response. It is consistent with materialism that experiences produce new mental representations, even if by necessity they represent things that are entailed by the old representations. We can use this to explain why Mary gains new abilities. We’ll start with a seemingly irrelevant example, one modelled on an example I got from Frank Jackson, via Derk Pereboom.

MOPs and Abilities
Billy and Suzy inhabit a world somewhat like this one. Their world has deterministic laws, which ours presumably does not. But the laws are in many cases chaotic, just as ours are. So in their world meteorology is very hard, just like it is this world. Both Billy and Suzy are given a meteorological present. Billy is given a full model of the planet’s weather patterns, replete with all the laws that the weather systems follow, and the initial conditions specified to a high enough degree that the model determines the weather for the next ten years to a high degree of accuracy. Given that the laws are chaotic, this means that he has to be told the initial conditions very precisely. Suzy, on the other hand, is given a full and accurate weather forecast for twenty major cities (including Ithaca) for each day for the next ten years.

In one sense Suzy’s information is a subset of Billy’s. Every world in which Suzy’s information is correct is a world in which Billy’s information is correct as well. But Billy’s information rules out some worlds that are compatible with Suzy’s information. So in that sense Billy has strictly more information about the world’s weather. But in another sense their information does not overlap. Suzy can easily tell you whether you should take an umbrella on any given day in Ithaca. For Billy to do that, he has to compute what the weather will be like in Ithaca given the initial conditions and the laws. And, we may well imagine, there is no computer in the world powerful enough to do this. (Remember that if Billy ever uses approximations in his computations, then the chaotic nature of the system will mean that his calculations will not be particularly accurate. So Billy’s calculations will be very difficult to do.)

Does Billy know what the weather will be in Ithaca next week? Different people may use the word ‘know’ differently in this case. On the one hand, he has information that determines what the weather will be. On the other hand, he lacks the ability to compute these mathematical consequences of his information. So it seems more natural to say he does not know. In this sense, Mary may not know what it is like to see red, even though this is a consequence of things she does know. (The Mary example, that is, requires some extreme computational idealisation as well as some epistemological idealisation if it is to be surprising that Mary comes to know something on leaving the room.) To avoid these complications, we’ll say that Billy has information about what the weather in Ithaca will be like, and if materialism is true, Mary has information about what it’s like to see red. Since ‘has information about’ is a technical enough term, we can stipulate its meaning, and we’ll stipulate that if a person’s information entails that \( p \), they have the information that \( p \), even if they are not capable of discovering that that entailment holds.

So Billy has all the information that Suzy has, though he represents it in a very different way. If it will be sunny and warm on day \( d \) in Ithaca, Suzy represents that by having the claim that it will be sunny and warm in her ‘belief box’. Billy represents this, like he represents everything, in the perfect meteorological model that he was given. This representational difference grounds a difference in abilities. Suzy, but not Billy, can tell you whether to take an umbrella to work. If Billy represented the information that he has the way Suzy represents it, then he too would have the ability to give you umbrella advice. If Billy builds a fast enough computer
to make real-time predictions of the weather, what he’ll get from the computer won’t be new information. Learning what the predictions of the model are for day $d$ doesn’t let Billy rule out any possible world that was previously consistent with his data. But it does give him a perspicuous representation of part of that information, and with that he has the ability to give meteorological advice.

The Abilities we get from Qualia

One of the themes that arises from time to time in the literature on qualia is that they are simple. This is something that Lewis discusses in “Should a Materialist Believe in Qualia?” for instance. Materialists think that the qualia are relatively complex neural entities or properties (or perhaps complex functional properties realised by complex neural entities or properties), and this may be taken to be a problem with materialism. But materialists can say something about the simplicity of qualia. Some qualia, at least the ones that are characteristically regarded as simple, are simple representations of complex states. A simple representation may be a very complex entity. A web site, for instance, is a very complex metaphysical thing, consisting of properties of a server combined with properties of the network that makes the files on the server into a web page. But if that web page just says “Billy kissed Suzy”, then it will be a fairly simple representation, although it is a complex metaphysical entity. (And although a kiss too is a complex entity, at least on the microphysical level.) Materialists can hold that qualia are likewise simple representations that are themselves complex and that represent (perhaps the same) complexities.

Now simple representations have various costs and benefits. The principle benefit is that they are easy to use in recognition and imagination. The principle cost is that they are less than completely useful in thinking about variations to the actual situation. We can illustrate this with a couple of cases, and then apply it back to qualia.

Alice is a normal person with a normal person’s ability to recognise when a violin is playing. She knows what things are like when a violin is playing, as we might put it. (I’m not imagining that Alice is perfect; perhaps she can’t tell a violin from related string instruments, but she can tell a violin from a trumpet, like most of us can.) Bob is not a normal person. When sound waves happen around him, he uses a recording device he keeps with him at all times to produce a digital image of those sound waves. He knows an algorithm that lets him compute (slowly and imperfectly) whether that sound wave is the sound of, inter alia, a violin. Now there is no information, in the salient sense, that Alice gets but Bob lacks. But Alice has a simpler way of representing that information. She can represent it simply as a violin playing. Until and unless he pulls out the algorithm and does the computations, Bob can only represent it as a mathematically defined wave.

Bob is obviously a strange creature, though not obviously stranger than Mary. But I think we can make the following two predictions about Alice and Bob. On the one hand, Alice will have greater imaginative powers than Bob. Ask Alice to imagine a violin playing, and she’ll be
able to do a tolerable job. Ask Bob to do the same, and unless he’s been thinking of a wave that he knows is the sound wave of a violin, he won’t obviously be able to do this. On the other hand, Alice won’t be generally reliable at saying what the interaction effects would be if some other sound (a trumpet, another violin, or a sine wave producer) were playing alongside the violin, and Bob will know more about this.

We can do the same thing with vision. Carla is a normal person who can (imperfectly) recognise cats. David is not a normal person. When he takes in a scene visually, he mentally represents it the same way that a digital camera represents the scene. If he wants to know whether a cat is in the scene, he uses a (slow and imperfect) algorithm for computing whether a digital image includes an image of a cat.

Again David is a strange case, but no stranger than Mary. And I think we can make predictions about him. Unless he has been considering some image of a cat recently, he won’t be able to imagine a cat, unlike Carla. And he is rather slow at figuring out whether there is a cat nearby. That’s enough, I think, to deny that he really knows what it’s like for there to be a cat nearby. On the other hand, David is better than Carla at saying what things would look like if there were some changes to the scenery, say that the light were greener, or dimmer, or if a strange checkerboard lighting pattern were put onto the visual image.

The people who use simple representations, Alice and Carla, seem to have greater imaginative capacities, as well as being able to quickly tell the coarse properties of what is around them. Those capacities seem to be crucial to our saying that Alice and Carla know what things are like when there is a violin playing or a cat nearby. And note that in this we are not talking about anything phenomenal. There being a violin playing or a cat nearby are paradigmatically physical facts about the world, not phenomenal facts about Alice or Carla.

If qualia are simple representations, then we’d predict that people who have the qualia, i.e. acquire the representational concepts, have (a) the ability to imagine situations in which things are as that representation represents them as, (b) know what it’s like for things to be that way, but (c) can’t tell how things would be if things were somewhat different. I think that’s basically the way that qualia do behave. They produce imaginative abilities, someone who has tasted Vegemite can imagine this, but they don’t produce much counterfactual knowledge. It’s frequently hard to tell, for instance, how things will taste when mixed together just on the basis of knowledge of how they taste separately.

At risk of affirming the consequent then, it looks like we have a decent chance of developing an inference to the best explanation argument. Qualia behave exactly as we’d expect were they simple representations of complex underlying states. By introspection we can tell that qualia are in some ways simple. And we have independent metaphysical reason to believe that if qualia represent anything, what they represent are not simple states. So I’m inclined to believe that they are simple representations of complex underlying states. So it isn’t surprising that they improve our imaginative abilities, even though what they represent is information that we could
independently acquire, and indeed that Mary has independently acquired. That’s just what simple representations do.

*Why We Think Mary Gains Information*

There are two ways that we might respond to the intuition that Mary learns something, or more precisely gains information, when she leaves the room. The first is to explain why it seems to Mary that she gains information. The other is why it seems to us that Mary would gain information. Both explanations are error theories; the first attributes an error to Mary, the second to us.

The New Mode of Presentation response is a response of the first kind. Mary erroneously, but reasonably, believes that she gains new information. As noted above, the problem with this kind of response is that it isn’t clear why Mary would make this error, since she knows what qualia represent.

The Ability Hypothesis is, I think, best thought of as the second kind of response. Mary does not make any errors when she leaves the room. We think, mistakenly, that she’ll gain knowledge. We think that she’ll gain information because we think, truly, that she’ll gain an ability and we know that in general it’s impossible to gain an ability without gaining some propositional knowledge. This last claim is true, this is not one of the general cases. In rare cases, like Mary’s, or like some of the cases I described above, the ability is gained without any new propositional knowledge. But this is rare, so it isn’t surprising that we go wrong.

Mary’s case in fact is rare twice over. It is rare to gain abilities without getting any propositional knowledge. The example of surgically restored Rickey I gave above might be one such case, but they are rare. It is particularly rare to gain knowledge of what something (be it the external world or one’s mind) is like without gaining propositional knowledge. If I know very little about cats, I won’t know much about what it’s like to have cats around. In practice, this lack of knowledge is why I don’t know what it’s like to have cats around. But this kind of ‘knowledge’ requires an ability as well as some propositional knowledge, and though its rare to have the propositional knowledge without the ability, Mary is in just that rare situation.