The knowledge argument is an argument about knowledge[1]

Tim Crane

1. Introduction

The knowledge argument is something that is both an ideal for philosophy and yet surprisingly rare: a simple, valid argument for an interesting and important conclusion, with plausible premises. From a compelling thought-experiment and a few apparently innocuous assumptions, the argument seems to give us the conclusion, a priori, that physicalism is false. Given the apparent power of this apparently simple argument, it is not surprising that philosophers have worried over the argument and its proper diagnosis: physicalists have disputed its validity, or soundness or both; in response, non-physicalists have attempted to reformulate the argument to show its real anti-physicalist lesson.

I disagree with both groups of philosophers: I think the argument is sound, but that it does not show that physicalism is false. What the argument shows is that there is some propositional or factual knowledge which you can only have if you have certain experiences. This is an important, interesting and fairly controversial conclusion, but it is consistent with both physicalism and dualism. The knowledge argument, I claim, is an argument about knowledge, not about the metaphysics of the mind.

I first expound the knowledge argument in its least contentious, controversial version, and give a little historical background. I then show why the standard (physicalist) critiques of the argument miss their mark, why the argument does not establish dualism, what the real lesson of the argument is, and why the dualists' required developments of the argument lack suasive or dialectical force. But first, the argument itself.

---

[1] This paper develops and corrects some of the ideas about the knowledge argument first put forward in Crane (2001 chapter 3), and Crane (2003). Thanks to Sam Coleman, Kati Farkas, Lizzie Fricker, Philip Goff, Henry Taylor and especially Howard Robinson for discussion of this argument, and to participants at a 2016 workshop in Cambridge on Robinson's latest book (Robinson 2016), and at the 2017 Midwest Epistemology Workshop in St Louis. The paper was written with the help of a grant from the John Templeton Foundation, *New Directions in the Study of the Mind*.

2. The knowledge argument summarised

Many things have been called the knowledge argument, but the essence of the argument is a thought-experiment where someone is imagined to have complete knowledge of a certain kind A, but lacks knowledge of another kind B. Scenarios are then sometimes envisaged in which they gain genuine knowledge of kind B, or it is tacitly assumed that they have it anyway. So the knowledge they had of the kind A cannot be all there is to know. Of course, the case which we are interested in is where A knowledge is 'physical' and B knowledge is 'phenomenal'; and the conclusion of interest is the one where the fact that not all knowledge is 'physical' is supposed to entail that physicalism is false.

The scenario envisaged by Frank Jackson (1982) is the thought-experiment of Mary the omniscient scientist who lives in a black-and-white room, and then sees something red for the first time. Nothing of great significance depends on the specific details of this particular version of the thought experiment. For example, if someone finds the scenario of someone learning physics in Mary's predicament hard to imagine — maybe because her physics books cannot all be in black and white — then they should imagine instead Mary being blind, and then recovering her sight. Physics can be in braille (Maddox 2007). By the same token, the argument does not depend on anything specific to vision — it could be formulated in terms of the knowledge given in hearing, smell or taste.

Given the basic assumptions of the thought experiment, we can express the argument in terms of the following three premises and a conclusion:

Premise 1: Mary knows all the physical facts about seeing red in the black and white room.

Premise 2: Mary learns something new when she leaves the room and sees red for the first time.

Premise 3: What Mary learns is a fact.

Conclusion: Not all facts are physical facts.

If physicalism is the doctrine that not all facts are physical facts, then the conclusion is the negation of physicalism.

A few brief clarifications about the dialectical role of the premises. The first premise is simply a stipulation of one of the features of the thought-experiment. Accepting this premise and accepting the thought-experiment are basically the same thing. Of course, one might reject the thought-experiment on broadly methodological grounds, as Daniel Dennett (1991) famously did; if one does this, there is no need to discuss the rest of the argument. After all, one can hardly say 'oh yes the thought experiment is fine, it's just that Mary wouldn't know *all* the physical facts in that situation' — for the thought experiment is, by definition, a supposedly possible situation in which Mary knows all the physical facts.

The second premise, unlike the first, is not simply a stipulation involved in the thought-experiment; it is rather something we are invited to conclude after being told the story of Mary leaving the room and seeing red for the first time. It would be a coherent reaction to the argument to accept the first premise and deny the second, for example; or to put it another way, to accept the coherence or intelligibility of the thought-experiment and yet reject premise 2. In this way, premise 2 is a distinct claim from the mere coherence of the thought-experiment.

Premise 3 makes explicit what is needed in order for the argument to be valid. Given the widely held view that there are at least three kinds of knowledge — knowing that, knowing how and knowing things — it could be claimed that the argument equivocates if it just involves premises 1 and 2. For it may be that premise 1 is about knowing that, but for all the thought-experiment says, premise 2 might be about knowing how (or 'ability' knowledge) or knowing things ('acquaintance' knowledge). The point of premise 3 is to explicitly rule out

those options. So those who adopt the ability response (Lewis 1983, Nemirow 1990, Mellor 1992) or the acquaintance response (Churchland 1987, Conee 2004, Tye 2009) can accept premise 2 and reject premise 3.

The story of Mary and the black-and-white room comes, of course, from Frank Jackson's articles 'Epiphenomenal Qualia' (1982) and 'What Mary did not Know' (1986). Because of the catchiness of the Mary story, the knowledge argument has sometimes been attributed to Jackson, although he himself has been generous in acknowledging influences and independent presentations of the argument. Of these presentations, attention must be drawn to Howard Robinson's concise statement at the opening of his *Matter and Sense*, published in the same year as Jackson's 'Epiphenomenal Qualia':

> Imagine that a deaf scientist should become the world's leading expert on the neurology of hearing. Thus, if we suppose neurology to be more advanced than present, we can imagine that he knows everything there is to know about the physical processes involved in hearing, from the ear-drum in. It remains intuitively obvious that there is something which this scientist will not know, namely *what it is like* to hear. (Robinson 1982: 4)

In his recent reflections on the knowledge argument, Robinson comments that he 'did not then treat this as a refutation of physicalism, but rather as a way of setting up the problem that faced the physicalist' (Robinson 2016). But it is easy to see how this brief vignette contains almost the entire argument as represented above — all it lacks is the claim that knowledge of what it is like to hear is knowledge of a fact. With this added, Robinson's argument is as clear a statement of the knowledge argument as any.

As a number of writers (e.g. Nida-Rümelin 2009) have pointed out, something like the knowledge argument is present in C.D. Broad's *The Mind and its Place in Nature*

(1927), Herbert Feigl's 'The "Mental" and the "Physical"' (1958), and Thomas Nagel's 'What is it Like to be a Bat?' (1974). To these precursors of the argument, I would add my own favourite statement of its basic idea, by Bertrand Russell:

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

However, the relationship between these precursors and the arguments of Jackson and Robinson is not straightforward. Robinson and Jackson in 1982 were opposing physicalism, as was Broad (in the name of 'mechanism'); but Russell draws no explicit conclusion about physicalism. Nor did Feigl and Nagel (at least in 1974) take their own 'knowledge arguments' to tell against the truth of physicalism. Feigl was defending a version of the identity theory, combined with the view that there are two kinds of knowledge of the mind-brain. Nagel's conclusion was that even though physicalism (materialism) is true, it is in a certain sense unintelligible to us.

The arguments of Jackson and Robinson, therefore, do add something new and clear which was not explicitly there in these predecessors: the use of these considerations about knowledge to undermine the doctrine of physicalism. Physicalists quickly rose to the challenge and over the past few decades have offered various criticisms of the argument. These have been effectively discussed in great detail in many places, and my aim here is not to present a full survey of these discussions. Rather, what I want to do is to identify what seem to me to be the essential features of the most common physicalist responses.

3. The usual physicalist responses

Making premise 3 explicit enables us to treat physicalists as objecting to the argument as unsound, rather than invalid. This useful simplification allows us to group physicalist responses to the argument into those which deny premise 2 or those which deny premise 3. Less common, in my experience, is the response mentioned above which denies the coherence of the thought-experiment — in other words, which denies premise 1. Here I will ignore this less popular response. It may be that it deserves a detailed treatment, but I will not give this here.

The denial of premise 3 can be dealt with quickly here; what a physicalist should say about premise 2 will take a little more unravelling. Those who deny premise 3 do so because they think either that Mary acquires only ability knowledge, or that she acquires only acquaintance knowledge. The existence and nature of ability knowledge and acquaintance knowledge has been intensely discussed in the last decade (e.g. Bengson and Moffett 2011; Tye 2009). It is true that if there were no such thing as ability knowledge or acquaintance knowledge, then these responses to premise 3 would fail. But this does not mean that if there were such things, the responses would succeed. To deny premise 3 you have to deny that what Mary learns is a fact; but it is plain that acquiring ability knowledge or acquaintance knowledge is compatible with learning a fact. What these physicalists have to show, then, is that Mary *only* acquires ability knowledge or acquaintance knowledge, and no propositional knowledge. There is a lot that can be said here, but I will rest with the significant observation that most physicalists make no explicit attempt do this. An exception is David Lewis, who bases his critique of the argument on the assumption that ‘phenomenal information’ (i.e. the objects of propositional knowledge of phenomenal states) must be rejected by physicalists (Lewis 1999: 285). I will argue below that physicalists can accept this kind of information — at least in one sense of the word — with impunity.

This brings us to premise 2. On the face of it, it is hard to deny that Mary learns something, as the story is usually told. In general, it seems that having new experiences gives us new knowledge — at the very least, it gives us some knowledge of what it is like, in a perfectly ordinary sense, to have such experiences. How could the experience of seeing red for the first time fail to give knowledge? One could, of course, reject the coherence of the whole scenario, but in the way I am setting the argument up, this is a denial of premise 1, not premise 2.

In his more recent guise as a physicalist, Frank Jackson has defended a denial of premise 2 which deserves mention, not least because of Jackson's role in setting up the original argument. Jackson's way of denying (2) appeals to the representational theory of experience: the phenomenal character of experience is exhausted by its representational content. So when Mary experiences something red she comes to be in a state which represents a property — phenomenal red — which does not exist:

> Physicalists can allow that people are sometimes in states that represent that things have a nonphysical property. Examples are people who believe that there are fairies. What physicalists must deny is that such properties are instantiated. (Jackson 2003)

Jackson's response to the knowledge argument would make sense if the argument assumed or entailed that what Mary learns about are 'qualia' conceived of as non-physical, non-intentional property. Then his response would be: Mary's experience is a representation of something which does not exist, a non-physical quale. But there are no such things as qualia. Therefore Mary's experience is an illusion and cannot be the basis of any new knowledge of the world. Mary learns nothing new.

Of course every theory of mind must allow that some conscious experiences are illusions. But the experience of colour is just one kind of conscious experience, and nothing

in the general structure of the knowledge argument requires either a realist theory of colours, or a commitment to qualia (conceived of as non-intentional properties of experience). The knowledge argument is compatible with a Galilean theory of colours, and also with intentionalism or representationalism about experience. For a Galilean about colour, the argument would just need to be formulated in terms of some experienced phenomenal property other than colour (taste, smell, pain…). And the argument only needs one case to make its point. So Jackson's response will only work in general if all these kinds of features of conscious experience are illusions. But this is barely credible. It is barely credible, for example, that pain is an illusion. But Jackson needs to make this claim if his argument is to work.

There is little to be gained for a physicalist, I think, by digging in their heels and simply insisting that Mary gains no knowledge. It's not that this position is incoherent, but rather that it lacks any plausibility, given our normal views about the relationship between knowledge and experience. And, as we shall see, it is completely unnecessary for a physicalist to insist on this, when a far more plausible response to the argument is available.

This far more plausible response is simple, and has been around in the literature at least since Terence Horgan's 1984 response to Jackson's original paper. The essence of it is this: the fact that Mary gains new knowledge does not in itself show that it is knowledge of something non-physical. In Mary's case, as in many everyday cases, we learn something new about the world even though the things we are learning facts about are things we already knew about in some other way. When we are experientially representing a thing or property which we have previously represented in some other way, this does not mean one has no new knowledge of this thing or property. This point has sometimes been explicated in terms of the idea of the intensionality of (propositional) knowledge (Chalmers 1996: 141), and sometimes in terms of the comparison with Frege's famous discussion of

the Morning Star and the Evening Star (Frege 1892). Two modes of presentation of something does not imply two things presented.

This response is clearly a way of *accepting* premise 2, since it acknowledges that Mary learns something new. So it would muddy the waters to describe the response as saying that Mary does not *really* learn anything new, but only learns 'an old fact in a new way' (I take the phrase from Chalmers 'Phenomenal Concepts and the Knowledge Argument' §1; in her useful survey, Nida-Rümelin (2009) calls this the 'Old Fact/New Knowledge' response). If Mary does learn a new fact, then it cannot be the same fact as something she already knew. Should we say that the ancient astronomers who discovered that Hesperus is Phosphorus merely learned that Hesperus is Hesperus in a new way? Of course not: learning that Hesperus is Phosphorus is in no sense whatsoever a 'way' of learning that Hesperus is Hesperus. The 'old fact in a new way' talk is a convoluted formulation of a much simpler idea: that Mary learns something new and this does not imply that the entities about which she gains this knowledge are distinct from the entities she knew all about in the black and white room.

It may be replied to this that, in philosophy at least, 'fact' is ambiguous: facts can be constituents of reality ('the world is the totality of facts': Wittgenstein 1921) or they can be objects of knowledge ('a fact is a thought which is true': Frege 1918-19). So perhaps premise 3 is ambiguous and the 'old fact in a new way' talk is just supposed to make this explicit: Mary learns a new (Frege-style) fact which is a mode of presentation of an old (Wittgenstein-style) fact.

It's true of course that these two notions of fact have been used in 20th century philosophy, and they are both perfectly legitimate notions which have their different uses. But this does not mean that the knowledge argument, as I stated it above, equivocates. The argument talks about *learning* facts and *knowing* facts — where the knowledge in question is clearly propositional. But it is only facts in Frege's sense which can be the

objects of propositional knowledge: what you know is true, it is something that can be learned, and something that can be conveyed to others. Being true, learned and conveyed to others — these are not features of facts in the Wittgensteinian 'constituent-of-reality' sense. Constituents of reality are not true, and you cannot learn or convey them (as opposed to learning or conveying truths about them). There is a sense in which you can know constituents of reality — you can know people, for example — but this is not propositional knowledge, since propositional knowledge is knowledge of truths and truths are not constituents of reality in the relevant sense. And the knowledge argument is explicitly about propositional knowledge. (The debate about the existence of propositions is not relevant to this point: no-one should think that whether physicalism is true turns on whether propositions exist.)

For this reason, then, it is a mistake to say that the argument equivocates: it's perfectly clear which notion of fact is involved in the argument, and the physicalist should have no objection to this notion of fact, so long as they accept the idea of objects (or contents) of propositional knowledge.

Many of those who respond to the argument in this way have put it in terms of Mary's gaining a new kind of concept, a 'phenomenal concept', which she did not have in the black and white room (Balog 2009; Papineau 2002). If concepts are in some way the constituents of states of knowledge, then this new concept would indeed explain why the knowledge is also new. However, phenomenal concepts are somewhat controversial (Crane 2005; Sundström 2011), and it would be better not to rest the defence of premise 2 on such controversial ideas. After all, the truth of premise 2 seems more obvious than any complex theoretical claims about concepts. And maybe Edouard Machery (2009) is right, and the idea of a concept will play no role in a future science of the mind. But this should not undermine the idea that Mary would gain new knowledge in the scenario described.

Fortunately, it is not necessary to adopt any novel theory of concepts in order to maintain premise 2. Mary could simply employ a demonstrative concept — '*that* is what red looks like!' — which is a kind of concept she also could employ when in the black and white room. A new experience can provide the opportunity for new knowledge — knowledge of a new truth or proposition — using concepts one had before. Just as one might express one's new knowledge of a person by saying 'That is the same person I met in Albuquerque last year', so Mary can use a demonstrative to express her new knowledge deriving from her new experience (Crane 2003). It is a further step, and not obligatory, to explain the truth of this demonstrative judgement in terms of phenomenal concepts in Balog's or Papineau's senses.

Howard Robinson (2016) has combined the idea that Mary gains factual knowledge with the idea that she gains knowledge by acquaintance. In laying out the options in response to the knowledge argument, Robinson says that 'Mary lacked and later acquired some factual knowledge concerning the nature of phenomenal colour' (2016: 21). I agree this is the right response to the argument. But I disagree with his further gloss on this:

> The information in question will not be propositional, but be a form of knowledge by acquaintance, but it will still be factual information concerning the nature of colour and colour experience. (2016: 21)

I see no difference between factual knowledge and propositional knowledge — factual knowledge is knowledge of Frege-facts or true propositions. I don't think there is another viable sense of factual knowledge, though there is a viable sense of 'fact': the Wittgensteinian sense. But for the reasons given above, Wittgensteinian facts are not the objects of factual knowledge. And knowledge by acquaintance — if it exists as a distinctive kind at all — is compatible with physicalism and its negation. So in his defence of the

knowledge argument Robinson should only appeal to factual knowledge. (I will return to Robinson's views in §5 below.)

The lesson I want to draw here is that the overwhelmingly plausible physicalist response to the argument — that the same things can be known about in different ways — is compatible with accepting all three premises of the argument. Those who make this response should not dispute the claim that Mary learns something new, and (if they want to maintain the analogy with other cases of the intensionality of knowledge ascriptions) they should not deny that Mary learns a fact. The standard response, then, should not dispute the premises of the argument as I have presented them above.

And since the conclusion follows from the premises, the standard response should not dispute the conclusion either. So defenders of this response should say that not all facts are physical facts. This might look like a strange thing for a physicalist to say, until we take into account what the knowledge argument means (and must mean, if it is to be intelligible) by 'physical fact'. This is really the key to understanding the argument, as we shall see.

4. Physicalism and physical truths

As we have seen, the knowledge argument employs the notion of a fact as an object of knowledge, so spelled out literally the conclusion says that not all objects of knowledge are physical objects of knowledge. I am arguing that physicalism should be unworried by this conclusion: someone can be a physicalist and accept that not all objects of knowledge are physical. To defend this position requires answering two questions: what is physicalism? And second, what makes an object of knowledge physical in the relevant sense?

There has been an extensive debate about the content of physicalism over the last few decades (Melnyck 2003, Montero 2013, Ney 2008, Papineau 2001, Poland 1994, Stoljar 2010). Most of these details — e.g. what counts as physics, 'Hempel's dilemma', the precise statement of the causal closure principle etc. — need not concern us here. What should be

uncontroversial about physicalism, these days at least, is that it is a thesis about the world, about reality, about what there is. It is an ontological thesis. Physicalists might say, for example, that all objects and events are physical (sometimes called 'token physicalism') or they might say that all properties are physical (sometimes called 'type physicalism'). Or they might say that all states of affairs in D.M. Armstrong's sense ('facts' in the Wittgensteinian sense) are physical. Or they might say that everything is determined by its physical nature, or that everything supervenes on physical reality. However they think 'physical' should be defined, physicalists these days tend to treat physicalism as an ontological doctrine, a doctrine about what reality contains.

It wasn't always like this. In his essay, 'Psychology in Physical Language' Rudolf Carnap described the thesis of physicalism as 'physical language is a universal language, that is, a language into which every sentence may be translated' (Carnap 1932-33: 107). It was characteristic of the logical empiricist philosophy of the day (and the later philosophy which it influenced) to formulate ontological doctrines in linguistic terms. This practice survived into, for example, Chisholm's (1957) attempts to find linguistic criteria of intentionality, which is supposed to distinguish the mental from the physical (see also Dennett 1969, chapter 1). But these days physicalism is not formulated as a doctrine about sentences.

The upshot is this. Given that there can be genuinely different facts in the 'object of knowledge' sense (Frege facts), without this difference corresponding to any ontological difference, it follows that physicalism as an ontological thesis is not a thesis about facts in the 'object of knowledge' sense. So the pre-physicalist Jackson was therefore quite wrong when he said that 'if physicalism is true, Mary knows all there is to know' (1986: 291). Physicalism can be true and yet Mary in the room can be ignorant of certain facts. The knowledge argument as stated in section 2 above does not refute physicalism.

This brings us to the second question: what makes an object of knowledge (or true proposition) 'physical' in the sense employed by the knowledge argument? This is not the same question as what 'physical' means for physicalists. For that latter question is about how to make an ontological classification; but the former question is about how to classify objects of propositional knowledge. The 'physical facts' according to the knowledge argument are all those true propositions that Mary could learn within the black-and-white room scenario. So let's ask what kinds of propositions Mary *could* learn within this scenario. Obviously she can learn the truths of physics. But what else? The pre-physicalist Jackson says that what Mary learns is knowledge that is part of physics, 'in a wide sense of "physical" that includes everything in completed physics, chemistry and neurophysiology' (Jackson 1986: 291). He is surely right that it clearly makes no difference to the story whether Mary learns facts about the physiology of the brain in addition to facts about fundamental physics. But the same could be said about the facts of theoretical psychology. It's in the spirit of the knowledge argument to say that Mary could learn everything in a completed scientific psychology in the room. And she would still not know what it was like to see red. Similarly, a blind person could learn a 'completed' psychology of vision and not know what it was like to see.

Pushing this idea a bit further, let's suppose that some kind of dualism is true, but it is a scientific dualism: a dualism which appeals to irreducible psychological laws which talk about or quantify over irreducible mental properties, which do not necessarily supervene on physical properties. Not all forms of dualism are like this of course; some forms of dualism are explicit that there cannot be a science of the mental at all. But it serves my purpose if there merely *could* be a form of dualism like this — indeed, Chalmers (1996) speculates about such a naturalistic dualism. Now if Mary were to learn such a theory in the black and white room, would it help her to know what it was like to see red? In so far as we have a grip on what this science might be, the answer seems to me clearly no.

What lesson should we draw from this? What do all these propositions have in common? It is misleading to say that anything that Mary can learn inside the room is physical, given that this word has an independent sense, the sense employed in the ontological debates about physicalism. So let's not introduce a second sense of 'physical' to go along with the second sense of 'fact'. What we should say instead is that the facts that Mary can learn in the back-and-white room scenario are the kind of facts that do not require any specific kind of experience. You may need some kind of experience in general in order to learn the full scientific theory of colour vision — you have to get the information somehow — but it is plausible that you don't need full chromatic colour vision. Similarly, you may need some kind of experience or other to learn the full scientific theory of taste and olfaction; but you don't need the experience of all the tastes. This is the point Russell is making when he says that a blind man can learn the whole of physics.

In an earlier paper (Crane 2003), I called the kind of knowledge Mary can acquire in the black-and-white room 'book-learning'. This was inspired by David Lewis's remark that 'intuitive starting point wasn't just that *physics* lessons couldn't help the inexperienced to know what it is like. It was that *lessons* couldn't help' (Lewis 1999: 281). The idea of something that can be learned in a book is vivid, but it is hard to make wholly explicit. What sorts of things can be learned in books, and what cannot? What the knowledge argument shows, at the very least, is that this is an important question for epistemology. The argument shows that the distinction between book-learning and non-book-learning is not the same as that between propositional and non-propositional knowledge, since Mary's new knowledge is, as I have argued, propositional. This is a significant result for epistemology.

John Perry (2001) has argued that Mary's new knowledge is just a special case of indexical knowledge. When Perry, following a trail of sugar in the supermarket in order to alert the person making the mess, discovers that the leak is coming from his own bag, he

gains the knowledge that *he* is making a mess. This is new propositional knowledge, but it is not book-learning: in order to acquire it, Perry had to recognise something about his position in the world at that moment. He had to occupy a specific position in the world; his knowledge was not available without occupying that position. Perry claims that Mary's position is comparable to this. And just as Perry the shopper's predicament has no ontological consequences, Mary's predicament does not either.

The analogy is plausible, but does it give an account the kind of knowledge Mary gains? I myself once thought something like this (Crane 2003), but now I think things cannot be that simple. Consider a Laplacean demon who has a complete theoretical, third-personal, objective knowledge of all the facts before Perry's shopper's discovery. The demon would be able to deduce that Perry's shopper will be able to know that *he* is the shopper making the mess. The demon would not, of course, be able to know that she herself, the Laplacean demon, was making a mess, because it is not true — she cannot truly think the proposition that Perry is thinking. But despite this, she knows — without remainder, I want to say — which proposition it is that Perry comes to know.

The situation is different when it comes to Mary's predicament. Given that she has total knowledge of physics, psychology, linguistics etc., the demon would be able to predict that Mary will think 'this is what red looks like' after she sees red for the first time. But if the demon herself had not seen red, then there would still be something significant lacking from her knowledge: what red looks like. This is a more substantial lack than in the Perry case. There the demon knew exactly what was going on, but in the case of seeing red, she was genuinely lacking something. For this reason, I don't think Perry is right that Mary's predicament is explained simply by the theory of indexicality. Something else is going on.

The conclusion of the knowledge argument is not simply that some knowledge requires experience; it is that there is some specific kind of knowledge which requires a specific kind of experience, and that this cannot be obtained in another way. Mary's

knowledge of what red looks like is knowledge of this kind. It is expressible in a proposition — *red looks like this* or *this is what red looks like* — but this proposition requires specific kind of visual experience in order to be learned. The idea that there might be knowledge of this kind is not trivial, but the knowledge argument provides one plausible reason for believing in it. Exactly how this species or kind of knowledge should be characterised, it seems to me, is a substantial question for epistemology.

5. How to get the dualist conclusion

In my interpretation of the argument, then, dualism does not follow from the conclusion of the knowledge argument. I disagree therefore with Robinson when he says, 'if the contrast between Mary and others, or between Mary before and after is a genuine one, then property dualism is established and one must adjust one's views accordingly' (Robinson 2016: 59). I reject this inference, not because I reject property dualism, but because one can accept that there is a genuine contrast, and still be a property monist. The contrast lies purely in Mary's experience and in her knowledge of the situation.

It is worth asking, then, why some philosophers think that the argument or something like it does establish dualism. In the final section of this paper I will venture a hypothesis about this, by way of a discussion of Robinson's version of the argument in his recent book, *From the Knowledge Argument to Mental Substance* (2016).

Robinson presents the argument in a slightly different way from the way I present it above. Here is his version:

> (1) Mary knows all those facts about the perception of chromatic colour which can in principle be expressed in the vocabulary of physical science.
>
> (2) Unlike those who have normal visual experiences, Mary does not know the phenomenal nature of chromatic colour (what it is like to perceive chromatic colour).

Therefore

(3) The phenomenal nature of chromatic colour in principle cannot be characterised using the vocabulary of physical science.

(4) The nature of any physical thing, state or property can be expressed in the vocabulary of physical science.

Therefore

(5) The phenomenal nature of chromatic colour is not a physical thing, state or property.

(Robinson 2016: 16-17)

One obvious difference between my version of the argument and Robinson's is that Robinson focuses on Mary's predicament inside the room, rather than on the change after the new experience. Robinson takes it for granted at this stage that Mary will gain new knowledge after seeing red, and of course I am happy to follow him in this.

What about the rest of his argument? Robinson's premise (1) is weaker than my premise (1) (it is a consequence of the latter) but otherwise his version introduces some somewhat different ideas. Premise (2) talks about knowing 'the phenomenal nature' of a property, which I take to be equivalent to knowing what it is like to experience that property. To show that (3) follows from (1) and (2) we need two things. First, we need to show that the knowledge in (1) is the same kind as that in (2) — i.e. propositional or factual knowledge — to avoid equivocation (as explained in §2 above). Robinson should adopt something like the reasoning I gave in §2 for the univocality of the knowledge claims here. And second, we should stipulate that 'expressed' and 'characterised' mean the same thing in this context. Given these two points, (3) will follow from (1) and (2). Premise (1) says Mary knows all the facts that can in principle be expressed by physics, and (2) says that

she does not know the facts about what it's like. So the facts about what it's like to experience the property cannot be something in principle expressed by physics.

(It would be possible to question the equation of 'expressed' and 'characterised', on the grounds that physics can characterise an experience without expressing it — a scientific description of a brain state, for a physicalist, can be a way of characterising something which is as a matter of fact an experience, but the scientific description does not in any plausible way 'express' the experience. However, I don't think Robinson is relying on any significant distinction between expressing and characterising, so I will not pursue this criticism of his argument.)

So far, Robinson's argument is fairly similar to the argument as I presented it above. But premise (4) is an additional premise. It says that the nature of a physical thing can be expressed in the vocabulary of physical science. This is unobjectionable in itself, but if the conclusion (5) is to follow, then premise (4) should say 'the *entire* nature of any physical thing can be expressed in the vocabulary of physical science'. If we do not add 'entire' (or some synonym), then it would be possible for a physicalist to say that physical science can express or characterise the nature of experiences (which are, as a matter of fact, identical with brain states), but that other aspects their nature can also be expressed using other descriptive materials. But this fact does not entail that this other aspect is not, ontologically speaking, a physical thing. On this view, although physical science can give a characterisation of the state, it would not give a full characterisation. Full characterisations can only be given when one employs all the concepts available; and some of these concepts are only available to those who have had the experience. But the experience can be a physical state for all that. In fact, this seems to me the essence of the 'phenomenal concept' response to the argument, stripped of the confused idea of an 'old fact in a new way'. I myself am sceptical about the specific idea of phenomenal concepts employed by

Papineau, Balog and others. But this is not relevant to the dialectical point here against Robinson.

It may be objected that once it is accepted that there are 'aspects' of things which cannot be described in physical terms, then physicalism has conceded the point. But this is not so; physicalists can allow such aspects, so long as they are conceived of epistemologically. The physicalism I have in mind has been helpfully labelled by Robert Howell 'inclusive subjective physicalism', according to which 'a complete physics will refer to every property and event that there is. There are simply ways of understanding those properties that will not be imparted by an understanding of the theoretical descriptions of physics' (Howell 2009: 316).

How can this kind of physicalist accept these aspects? Physicalists accept experiences, and (I would argue) they should accept that you don't know what it's like to have a kind of experience unless you have had one of that kind. So they should say that experiences have the following aspect, feature or property: *they are such that you cannot know what they are like without having had them*. Therefore you cannot know what they are like through book-learning alone (as Einstein is supposed to have said, 'science cannot give you the taste of chicken soup'). But this is just another way of distinguishing between the knowledge you get from books (whatever books are exactly) and the knowledge you get from tasting something. And that distinction itself is just a consequence of the sort of thing that tastes (etc.) are.

So Robinson's argument will get its dualist conclusion if premise (4) is modified to include the word 'entire': 'the entire nature of any physical thing, state or property can be expressed in the vocabulary of physical science'. But this is not something a physicalist has to accept. That was the upshot of my discussion in this section and §3.

Other versions of the knowledge employ the idea of knowing something in its entirety — but in the mirror image, so to speak, of the claim that Robinson needs. They

use instead the idea that experiencing something enables you to know its *phenomenal nature* in its entirety. Reflection on an experience therefore enables you to know, *ipso facto*, that it is not also a physical phenomenon. As Nida-Rümelin comments:

> The intuitive idea … has been expressed in different ways. Some say that qualia 'have no hidden sides'. Others say that qualia are not natural kind terms [*sic*.] in that it is not up to the sciences to tell us what having an experience of a particular kind amounts to (we know what it amounts to by having them and attending to the quality at issue). It is quite clear that an account of this intuitive idea has to be one of the ingredients of a dualist defense of the knowledge argument. (Nida-Rümelin 2015)

What Nida-Rümelin here calls the intuitive idea is not just one of the ingredients of a dualist knowledge argument — it is, arguably, *the* active ingredient in a wholly different argument from the one discussed above (§§2-3). For if you accept the idea that experience allows you to know an experienced property in its entirety, then you need little else to get the dualist conclusion. For it is plain that you cannot learn that an experience is a physical state merely by having that experience and reflecting upon it. So if the experience is supposed to give you knowledge of the entire nature of a phenomenal properties, you could refute the identity theory simply by reflecting on your experience.

Philip Goff has recently put this point by saying that concepts of phenomenal states are 'transparent' — they reveal the nature of those states (2017: 74). Although he himself thinks that phenomenal concepts are transparent, Goff argues plausibly that 'the knowledge argument does not have the resources to establish' this claim, 'without which Mary's knowledge is no threat to physicalism' (2017: 75). This is, in effect, one of the lessons of the present paper.

There are two substantial assumptions, then, that can be used to derive dualist conclusions when added to the knowledge argument's premises. One is the assumption Robinson needs: that the entire nature of any physical thing, state or property can be expressed in the vocabulary of physical science. The other is what Nida-Rümelin calls the 'intuitive idea', and what Goff calls 'transparency': that the entire nature of a phenomenal property can be known from experiencing it. The first assumption can be used when you concentrate, as Robinson does, on what Mary doesn't know in the black and white room; the second assumption can be used when you concentrate on what Mary comes to know in having the relevant new experience.

These are both very strong assumptions. And although it is possible to build arguments against physicalism using one or both of them (see Nida-Rümelin 2007; Goff 2017), it should be clear that they cannot be derived from the uncontroversial premises of the original 1982 knowledge arguments of Jackson and Robinson. But without these additional assumptions, physicalism is untouched by the argument.

6. Conclusion

In this paper, I have not disputed the significance of the knowledge argument itself, but only its usual interpretations. I have disputed both the dualist interpretation, and the usual physicalist interpretations. Although not a physicalist myself, I argued above that the argument is not effective against physicalism. But this is not because the argument itself, considered in terms of the most plausible reading of its premises and conclusion, is invalid or unsound. It is because the real target of the argument is not physicalism, but a certain conception of knowledge. So instead of an unsound argument against physicalism, we have a sound argument which identifies a particularly important form of knowledge. The precise nature of that knowledge is a matter for further epistemological investigation.

One final point. Physicalists and non-physicalists have commented on the fact the knowledge argument moves from epistemological premises to metaphysical conclusions, and some have wondered how this is possible. The answer is that in the case of what I think of as the core of the knowledge argument, it is not. Whether or not such metaphysical conclusions can be drawn from other epistemological premises I will not discuss; I restrict myself to the conclusion that the knowledge argument itself does not yield any significant metaphysical conclusions.

References

Balog, Katalin (2009) 'Phenomenal Concepts' In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *Oxford Handbook in the Philosophy of Mind* (Oxford: Oxford University Press) 292–312.

Bengson, John and Marc Moffett (eds.) (2011) *Knowing How: Essays on Knowledge, Mind, and Action* (Oxford: Oxford University Press).

Broad, C. D. (1926) *The Mind and Its Place in Nature* (London: Routledge and Kegan Paul).

Carnap, Rudolf(1932-33) 'Psychology in Physical Language' *Erkenntnis* 3: 107-42.

Chalmers, David (1996) *The Conscious Mind* (Oxford: Oxford University Press).

Chalmers, David (2004) 'The Representational Character of Experience' in B. Leiter (ed.) *The Future for Philosophy* (Oxford: Oxford University Press) 153–181.

Chalmers, David (2006), 'Perception and the Fall from Eden' in Tamar Szabó Gendler and John Hawthorne (eds.), *Perceptual Experience* (Oxford: Oxford University Press).

Chisholm, R.M. (1955–6) 'Sentences about Believing,' Proceedings of the Aristotelian Society 56: 125–148.

Churchland, P. M. (1997) ‘Knowing Qualia: A Reply to Jackson’ in N. Block, O. Flanagan, and G. Güzeldere (eds.) *The Nature of Consciousness* (Cambridge, MA: MIT Press) 571-578.

Conee, Earl (2004) ‘Phenomenal Knowledge’ in Yujin Nagasawa, Peter Ludlow & Daniel Stoljar (eds.), *There's Something About Mary* (Cambridge, MA: MIT Press).

Crane, Tim (2001) *Elements of Mind* (Oxford: Oxford University Press).

Crane, Tim (2003) ‘Subjective Facts’ in H. Lillehammer and G. Rodriguez-Pereyra (eds.) *Real Metaphysics* (London: Routledge) 68–83.

Crane, Tim (2005) ‘Papineau on Phenomenal Concepts’ *Philosophy and Phenomenological Research* 71: 155–162.

Dennett, Daniel C. (2007) ‘What RoboMary Knows’ in Torin Alter and Sven Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge* (Oxford: Oxford University Press) 15–31.

Dennett, Daniel C. (1969) *Content and Consciousness* (London: Routledge and Kegan Paul).

Dennett, Daniel C. (1988) ‘Quining Qualia’ in A. Marcel and E. Bisiach (eds.) *Consciousness in Contemporary Science* (Oxford: Oxford University Press) 42–77.

Dennett, Daniel C. (1991) *Consciousness Explained* (London: Allen Lane).

Feigl, Herbert (1958) ‘The “Mental” and the “Physical”’ in H. Feigl, M. Scriven, and G. Maxwell (eds.) *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press) 370-497.

Frege, Gottlob (1892) ‘On Sense and Reference’ in A.W. Moore (ed.) *Meaning and Reference* (Oxford: Oxford University Press 1993).

Frege, Gottlob (1918–1919) ‘Thoughts’ in N. Salmon and S. Soames (eds.) *Propositions and Attitudes* (Oxford: Oxford University Press 1988) 33–55.

Goff, Philip (2017) *Consciousness and Fundamental Reality* (Oxford: Oxford University Press).

Harman, Gilbert. (1990), 'The Intrinsic Quality of Experience' in J. Tomberlin (ed.) *Philosophical Perspectives* 4 (Atascadero: Ridgeview); reprinted in N. Block, O. Flanagan, and G. Guzeldere (eds.) *The Nature of Consciousness* (Cambridge, MA: MIT Press 1997) 663-676.

Horgan, Terence (1984) 'Jackson on Physical Information and Qualia' *Philosophical Quarterly* 34: 147–152.

Howell, Robert (2009) 'The Ontology of Subjective Physicalism' *Nous* 43: 315-345.

Jackson, Frank (1982) 'Epiphenomenal Qualia' *Philosophical Quarterly* 32: 127–136.

Jackson, Frank (1986) 'What Mary Did not Know' Journal of Philosophy 1986; reprinted in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness* (Cambridge, MA: MIT Press 1997) 567-570.

Jackson, Frank (1995) 'Postscript to "What Mary Did not Know"' in Paul Moser and J.ºD. Trout (eds.) *Contemporary Materialism* (London: Routledge) 192-198.

Jackson, Frank (2002) 'Representation and Experience' in H. Clapin, P. Slezack, and P. Staines (eds.) *Representation in Mind: New Approaches to Mental Representation* (Westport, CT: Praeger) 107–124.

Jackson, Frank (2003) 'Mind and Illusion' in Anthony O'Hear (ed.), *Minds and Persons*, (Royal Institute of Philosophy Supplement. Cambridge: Cambridge University Press) 421–442.

Levine, Joseph (1983) 'Materialism and Qualia: The Explanatory Gap' *Pacific Philosophical Quarterly* 64: 354–361.

Lewis, David (1999) 'What Experience Teaches' in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press) 262-290.

Machery, Edouard (2009) *Doing without Concepts* (Oxford: Oxford University Press).

Maddox, Steve (2007) ‘Mathematical Equations in Braille’ *MSOR Connections* 7, 45-48.

Mellor, D.H. (1992–1993) ‘Nothing Like Experience’ *Proceedings of the Aristotelian Society* 92: 1–16.

Melnyk, Andrew (2003) *A Physicalist Manifesto: Thoroughly Modern Materialism* (Cambridge: Cambridge University Press).

Montero, Barbara (2013) ‘Must Physicalism imply the supervenience of the mental on the physical?’, *Journal of Philosophy*, 110(2): 93–110.

Nagel, Thomas (1974), ‘What Is It Like to Be a Bat?’ *Philosophical Review* 83: 435–450.

Nemirow, Lawrence (1990) ‘Physicalism and the Cognitive Role of Acquaintance’ in W.ºG. Lycan (ed.) *Mind and Cognition* (Oxford: Blackwell) 447-461.

Ney, A., 2008, ‘Physicalism as an Attitude’, *Philosophical Studies*, 138: 1–15.

Nida-Rümelin, Martine (2007) ‘Grasping phenomenal properties’, in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism* (Oxford: Oxford University Press) 307–349.

Nida-Rümelin, Martine (2009), ‘Qualia: The Knowledge Argument’, *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2015/entries/qualia-knowledge/>.

Papineau, David (2001) ‘The Rise of Physicalism’ in Carl Gillett and Barry Loewer (eds.) Physicalism and Its Discontents (Cambridge: Cambridge University Press) 3-36.

Papineau, David (2002) *Thinking About Consciousness* (Oxford: Oxford University Press).

Perry, John (2001), *Knowledge Possibility and Consciousness* (Cambridge, MA: MIT Press).

Poland, Jeffrey (1994) *Physicalism: The Philosophical Foundations* (Oxford: Clarendon Press).

Robinson, Howard (1982) *Matter and Sense* (Cambridge: Cambridge University Press).

Robinson, Howard (2016) *From the Knowledge Argument to Mental Substance: Resurrecting the Mind* (Cambridge: Cambridge University Press).

Russell, Bertrand (1927) *The Analysis of Matter* (London: George Allen and Unwin).

Stoljar, Daniel (2010) *Physicalism* (London: Routledge).

Sündström, Per (2011) 'Phenomenal Concepts' *Philosophy Compass*, Vol. 6, no 4, 267-281.

Tye, Michael (2009) *Consciousness Revisited: Materialism without Phenomenal Concepts* (Cambridge, MA: MIT Press).