Introduction

Daniel Stoljar and Yujin Nagasawa

Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of ‘physical’ which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies. … It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning—she will not say “ho, hum.” Hence, physicalism is false. (Jackson 1986, p. 291, Chapter 2, this volume)

Frank Jackson’s knowledge argument is one of the most discussed, important, and controversial arguments in contemporary philosophy, and in the debates about consciousness to which philosophers make a contribution.¹ On its face, the argument purports to show that physicalism—the doctrine that everything is physical or supervenes on the physical—is false or at least is incompatible with the existence of consciousness. And it does so on the basis of easily understood premises and by employing an apparently impeccable argumentative structure.

Yet the argument has implications wider than simply the question of physicalism. For one thing, even those philosophers who think of physicalism as not up for serious debate regard the knowledge argument as zeroing in on that aspect of conscious experience that is most puzzling, and most challenging for their view. Moreover, for many the real lesson of the argument is not the falsity of physicalism, but rather the possibility that conscious experience lies quite beyond our comprehension, and perhaps beyond the limits of rational inquiry generally.

It is therefore important to achieve a deeper understanding of the argument. That is the impetus behind the present volume. Its purpose is to gather together the main papers in which Jackson has presented—and later rejected²—his argument, plus some classic and more recent papers in which the argument is discussed and debated.³ In turn,

¹ The knowledge argument is popular beyond the philosophy community. It plays a crucial role in David Lodge’s (2001) novel, Thinks... in which a cognitive scientist at the ‘University of Gloucester’, has an affair with a widowed novelist. The knowledge argument is also discussed in the recent series Brainspotting which showed on the UK on Channel 4 in 1996. Thanks to Martin Davies and Michael Smith for bringing these to our attention.

² For his reasons see §6 below, Foreword, and Chapters 18 and 19 of this volume.

³ The literature on the knowledge argument is voluminous. We have tried to include the key papers from each of the many responses to the argument. But inevitably there are fine papers we have not been able to include and there are papers that someone else putting together a similar volume would have preferred to some of those we have included.
the purpose of this introduction is to set out the main themes of the discussion, and explain the relations among them.

1. The Knowledge Intuition.

Many arguments in philosophy of mind, and in other areas of philosophy, are based on intuitions that are extremely persistent and widespread, intuitions which (it is natural to speculate) have their sources in deep aspects of human psychology. Perhaps the best-known such intuition is the inverted spectrum, the idea, as Locke put it, that “the same Object should produce in several Men’s Minds different Ideas at the same time; v.g. the Idea, that a Violet produces in one Man’s Mind by his Eyes, were the same that a Marigold produced in another Man’s and vice versa”. The inverted spectrum seems so compelling that one suspects that it is to some extent ‘latent’ in human thinking—that the contingent structure of the mind is such that we are prone to find the inverted spectrum possible (cf. Block 1990).

Something similar is plausible in the case of the knowledge argument. At the beginning the 1982 paper (Chapter 1, this volume), Jackson describes the intuition that he says lies behind the argument:

Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won’t have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise or seeing the sky. (p. 127)

The intuition that Jackson is pointing to here concerns knowledge: no amount of knowledge of a certain sort, a physical sort, no matter how it is put together, is going by itself to suffice for knowledge of a different sort, a phenomenal sort. This intuition is not yet the knowledge argument, but it is an intuition on which that argument is based. Let us call it the knowledge intuition.

The knowledge intuition is properly called an intuition in this sense: it presents us with a prima facie modal truth, a truth about what is possible. What it says is that it is possible for someone to know all the physical truths about, e.g. smelling a rose, without knowing what it is like to smell a rose. The issue is not about what in fact anybody does


5 The prominence in contemporary philosophy of the phrase ‘what it is like’ is owing to Nagel 1974, but as far as we are able to judge, the phrase was originally used, in the context of phenomenal consciousness, in Wittgenstein’s Remarks on the Philosophy of Psychology, Vol. 1, (University of Chicago Press, 1980) which dates from 1946-7. (A comparable usage is Farrell 1950). At §91, Wittgenstein writes: ‘The ‘content’ of experience, of experiencing: I know what tooth-aches are like, I am acquainted with them, "I know what it’s like to see red, green, blue, yellow, I know what it’s like to feel sorrow, hope, fear, joy, affection...". (It is interesting to note that the material following the asterisk was written by Wittgenstein in English rather than German.)
know; perhaps nobody knows all the physical truths about smelling roses. And the issue is not whether it is psychologically possible to know all the physical truths of this sort without knowing the phenomenal truths; perhaps this is not psychologically possible. The issue is rather logical: knowledge of the physical sort does not suffice logically for knowledge of the phenomenal sort.

If the knowledge intuition is something which is present in ordinary thought, it could scarcely be original with Jackson, and he certainly doesn’t say that it is. In fact, not only is it easy to find examples of the knowledge intuition in the literature prior to Jackson, it is also easy to find the intuition being discussed in a context that is quite distinct from the one in which he discusses it. To illustrate, consider the following passage from Bertrand Russell’s *The Problems of Philosophy*:

Physical science, more or less unconsciously, has drifted into the view that all natural phenomena ought to be reduced to motions. Light and heat and sound are all due to wave motions, which travel from the body emitting them to the person who sees light or feels heat or hears sound. That which has the wave-motion is either ether or ‘gross matter’ but in either case is what the philosopher would call matter. The only properties which science assigns to it are position in space, and the power of motion according to the laws of motion. Science does not deny that it may have other properties; but if so, such other properties are not useful to the man of science, and in no way assist him in explaining the phenomena.

It is sometimes said that ‘light is a form of wave-motion’, but this is misleading, for the light which we immediately see, which we know directly by means of our senses, is not a form of wave motion, but something quite different—something we all know if we are not blind, though we cannot describe it so as to convey our knowledge to a man who is blind. A wave-motion, on the contrary, could quite well be described to a blind man, since he can acquire a knowledge of space by the sense of touch; and he can experience a wave-motion by a sea voyage almost as well as we can. But this, which a blind man can understand, is not what we mean by light: we mean by light just that which a blind man can never understand, and which we can never describe to him. (Russell, 1912, p. 27-8).

Russell regards a certain sort of empiricism to be obvious, namely, that experience is necessary for the sort of phenomenal knowledge of light usually available to the sighted. It follows from this that a blind person—someone who by definition has not had the relevant experiences—cannot attain that sort of knowledge. On the other hand, Russell regards it as equally obvious that physical knowledge is in principle available to a blind person. Russell therefore endorses the knowledge intuition: no amount of knowledge of physical sort is by itself sufficient for phenomenal knowledge. Nevertheless, Russell is concerned here, not with metaphysics or physicalism, but with epistemology. In particular, he wants to argue that our epistemic position with respect to physical objects is analogous to the blind person’s epistemic position with respect to light: “physical objects remain unknown in their intrinsic nature, at least as far as discovered by means of the senses” (1912, p. 34).
The knowledge intuition is also present in discussion of a question that is related to Russell’s but different from it, the question in philosophy of perception of what makes a psychological faculty one of sensory perception rather than simply a way of registering information about the world.\(^6\) An interesting example is provided by British engineer J. W. Dunne. Along with a rather odd theory of time and dreaming, Dunne provides the following thought experiment in his *An Experience with Time* (1927):\(^7\)

Let us suppose that you are entertaining a visitor from some country where the inhabitants are all born blind; and that you are trying to make your guest understand what you mean by ‘seeing’. You discover, we will further assume, that the pair of you have, fortunately, this much in common: You are both thoroughly conversant with the meaning of all the technical expressions employed in the physical sciences.

Using this ground of mutual understanding, you endeavour to explain your point. You describe how, in that little camera which we call the ‘eye’, certain electromagnetic waves radiating from a distant object are focused on to the retina, and there produce physical changes over the area affected; how these changes are associated with currents of ‘nervous energy’ in the criss-cross of nerves leading to the brain-centres, and how molecular or atomic changes at those centres suffice to provide the ‘seer’ with a registration of the distant object’s outline.

All this your visitor could appreciate perfectly.

Now, the point to be noticed is this. Here is a piece of knowledge concerning which the blind man had no previous conception. It is knowledge which he cannot, as you can, acquire for himself by the ordinary process of personal experiment. In substitution, you have offered him a *description*, framed in the language of physical science. And that substitute has served the purpose of conveying the knowledge in question from yourself to him.

But in ‘seeing’ there is, of course, a great deal more than mere registration of outline. There is, for example—colour. … Physical description cannot here provide the information which experience could have given. (pp. 13-15)

Dunne’s example is reminiscent both of Russell’s example in the passage quoted above, and also to H. G. Wells’ story, ‘The Country of Blind’ (1904), in which a sighted stranger tries to convince the blind locals that he has one more sense than them. (Jackson mentions the story in the 1982 paper).\(^8\) Dunne takes it for granted that you would need to be able to see to have the concept of seeing, and on that basis endorses (something like) the knowledge intuition. But since he is apparently primarily concerned with the nature of sight rather than the falsity of physicalism, we again have a use of the knowledge intuition that is different from Jackson’s.

---

\(^6\) This question was of course posed and discussed by Grice 1962.

\(^7\) We are indebted to John Bigelow for bringing this passage to our attention.

\(^8\) Indeed, discussion of the blind is standard in the literature from at least the 17th century. See the quotation from Locke in Van Gulick (*Chapter 16*, this volume).
Unlike Russell and Dunne, the bearing of the knowledge intuition on epistemology or philosophy of perception is not important for Jackson. What is important for him is its bearing on the metaphysical thesis of physicalism, the doctrine, as he puts it, that “all (correct) information is physical information” (1982, p. 127). But what exactly is the connection between the knowledge intuition and physicalism?

2. From Intuition to Argument.
In the 1982 paper, Jackson notes the simplest way in which the connection between the knowledge intuition and physicalism might be forged:

Nothing you could tell me of a physical sort captures the smell of a rose, for instance. Therefore, Physicalism is false. (p. 127)

He continues:

By [my] lights this is a perfectly good argument. It is obviously not to the point to question its validity, and the premise is intuitively obviously true…to me. I must, however, admit that it is weak from a polemical point of view. There are…many who do not find the premise intuitively obvious. The task then is to present an argument whose premises are obvious to all, or at least to as many as possible. (pp. 127-128)

Suppose we say that any argument whose premises constitute or contain the knowledge intuition, and whose conclusion is that physicalism is false, is properly called a knowledge argument. It would then appear that the simplest form of a knowledge argument is the one that Jackson presents in the first of these two passages. And, as he goes on to say in the second passage, the goal of the 1982 paper is to find a version of the argument which improves on the dialectical weakness of this simplest version. In a moment we will consider how the argument that Jackson develops differs from the simplest version that he begins with, and we will also take up the question of whether “it is obviously not to the point to question [the argument’s] validity.” (pp. 127-128). But first we should note that, if we define a knowledge argument in these terms, it is not hard to find examples of it in the literature.9

One example can be found in C. D. Broad’s Mind and Its Place in Nature (1926), and in particular in the chapter “Mechanism and Its Alternatives”.10 In that chapter, Broad is concerned with “the logical distinction between mechanism and emergence” (p. 70). The problem with mechanism as he conceives of it, is that “a mathematical archangel”—that is, a being who is both logically omniscient and who knows all the mechanistic truths about various chemical compounds—would know all the truths if mechanism is true. But it appears that the archangel would not know all the truths:

---

9 What follows are examples of the knowledge argument from the twentieth century. So far as we can judge, it is difficult to find clear examples of the argument prior to this, though of course there are many close cousins of the argument in the literature, e.g. Leibniz’s Mill Argument, and Descartes’ Conceivability Argument. For some informative historical discussion, see Güzeldere 1998 and Vinueza 2002.

10 Thanks to David Chalmers for bringing this passage to our attention.
Take any ordinary statement, such as we find in chemistry books; e.g. “Nitrogen and Hydrogen combine when an electric discharge is passed through a mixture of the two. The resulting compound contains three atoms of Hydrogen to one of Nitrogen; it is a gas readily soluble in water, and possessed of a pungent and characteristic smell.” If the mechanistic theory be true, the archangel could deduce from his knowledge of the microscopic structure of atoms all these facts but the last. He 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. If the existence of the so-called secondary qualities, or the fact of their appearance, depends on the microscopic movements and arrangements of material particles which do not have these qualities themselves, then the laws of this dependence are certainly of the emergent type. (pp.71-2).

In this passage, Broad is endorsing a version of a knowledge argument in the sense we have introduced: Broad takes the knowledge intuition as his premise, and draws from this premise the conclusion that physicalism (“mechanism”) is false.

A different presentation of the knowledge argument is in B. A. Farrell’s paper ‘Experience’ published in 1950:

Suppose we had obtained from our Martian visitor all the information that we, as psychologists and physiologists, could obtain about his sensory faculties. We should probably still want to say: “I wonder what it would be like to be a Martian—with his pseudo-radio sense, able, for example, to listen to whatever wave length he chooses. Extraordinary!” This seems to be a perfectly sensible remark. But if there was nothing more to be discovered about the Martian than his actual and possible responses, then this would not be a sensible remark. We would know what it would be like to be a Martian and there would be no point wondering about it. So there is something more to be learnt about the Martian, and that is what his experience is like. (1950, p. 183).

Farrell’s exposition is somewhat less straightforward than Broad’s; in part he is picking up on themes already present in the passages from Russell and Dunne. Yet it is obvious from his article that Farrell is interested in a line of thought that takes us from the knowledge intuition to the falsity of physicalism—a knowledge argument in our sense. His intriguing response is that this line of thought presupposes a mistaken conception of experience, a conception which forgets that experience is, as he puts it, featureless.  

———

11 What Farrell means by ‘featurelessness’ is very closely related to the phenomenon of the transparency of experience, much discussed by contemporary philosophers, including Jackson in Chapter 19, this volume.
Still further versions of the argument can be found in the debates over the identity theory in the 1950s particularly in the discussion between Paul E. Meehl and Herbert Feigl. Here is Feigl:

Let us assume that a complete explanation of animal and human behavior can be achieved by reduction to the basic physical laws, and that the structures (initial and boundary conditions) of organisms can be described in purely…[physical]…terms; then there is no need for the phenomenal terms—just as there would be no need for typically biological or physiological concepts. They would all be “reduced” to whatever are the concepts of the “ultimate” physics (e.g., something like the concepts of current atomic, quantum and field physics). The Martian’s repertory—if he has a repertory of qualities of immediate experience at all (i.e., if he is not a “mere robot”)—may not in any way overlap with that of us earthlings. In that case he would lack altogether any “acquaintance” with the qualities of our “raw feels”. He would consequently also lack the sort of “empathy” that humans can have for each other. The physicalist would formulate this, of course, by pointing to essential differences between the Martian and the human central states and processes. The Martian would thus not know “what colors look like”; “what musical tones sound like”. Nevertheless he would be able to explain—and possibly also predict—all of human behavior on the basis of his micro-theories. His theories may be expressed in a notation (reflecting concept formation) utterly different from our basic physics—but his physics would be completely translatable into ours and vice versa.

Now, the question arises: Is there something about human beings that the Martian does not (and never could) “know”? (1967, pp. 139-140)\(^\text{12}\)

Feigl himself goes on to answer this question in the negative—“I maintain that, given enough time and intelligence, the Martian … would in principle arrive at a complete explanation of the behavior of earthlings” (1967, p. 140)—but it is clear that in raising the question he thinks that a positive answer is at least possible, and that such a positive answer, if it could be defended, would defeat physicalism. In short, he thinks that the knowledge intuition is false, but that if it were true it would defeat physicalism.

Unlike Feigl, Meehl regards the knowledge intuition as undoubtedly true:

Suppose that knowers K1 and K2 share knowledge of the Utopian scientific network, including of course the psychophysiology of vision and the psycholinguistics of color language. However, K2 is congenitally blind and although he has recently undergone a corneal transplant, it is very shortly postsurgery and he has not as yet experienced any visual raw feels. Does K1 know anything K2 does not know?

I have put this question to a number of “plain men” as well as to some not so plain. The natural tendency is to say that K1 knows something that K2 does not know, to wit, “what red looks like”. Without taking vulgar speech or the

\(^{12}\) Original publication: 1956
theories it embodies as criteria of truth, I must confess that my own instincts in this matter are very like those of the plain man. (1966, p. 151)

However, while Meehl thinks the intuition is true, he does not think the inference from the knowledge intuition to the conclusion that physicalism is false is a reasonable one—a point noted in the last footnote of Jackson 1982. Meehl writes:

Some have felt that this line of thought is adverse to the identity theory. But I am unable to see why...Can anything be inferred about ontological identity from the admission that a knower is historically acquainted “by direct experience” with raw-feel quality Q can, by virtue of that acquaintance, recognize a new instance of it and token it appropriately, whereas a knower who lacks such acquaintance “by direct acquaintance” cannot do so? (pp. 157-158)

For Meehl, the knowledge intuition only shows something epistemological, i.e. that a person who has no experience of red does not understand or does not fully understand the meaning of the term ‘red’, or the concept of red. But he insists that it does not follow from this that anything ontological—like the denial of physicalism or the identity theory—is true. As we will see later on, Meehl’s suggestion here—that there is something illicit in the knowledge argument since it apparently moves from an epistemological premise to an ontological conclusion—is something that has been taken up by many of the contemporary critics of Jackson’s argument.

Indeed, in Broad, Farrell, Meehl and Feigl, we see four distinct responses to the argument, responses which are still very much alive today: Broad accepts the argument as sound; Farrell rejects the argument on the grounds that it rests on a mistaken view of experience; Feigl accepts the argument as valid, but insists that the premise—the knowledge intuition—is false; finally Meehl rejects the argument as invalid.\(^\text{13}\)

3. The Argument Itself.
If the knowledge intuition and the knowledge argument were widespread in philosophical literature prior to his papers, what is distinctive about Jackson’s contribution? What is the difference between the knowledge argument of (e.g.) Broad and the knowledge argument of Jackson?

3.1. The Thought Experiment
The first difference is that the thought experiment with which Jackson operates provides a much better illustration of the knowledge intuition than other thought experiments. As is of course very well-known, Jackson’s thought experiment introduces Mary, a famous neuroscientist. (A concise statement of this example is to be found in the passage we

\(^{13}\) Nicholas Maxwell (2000, 2001) suggests that he introduced the knowledge argument in 1960's in Maxwell (1966, 1968). However, it is not obvious that Maxwell's argument is a clear example of a knowledge argument in the sense we intend. Maxwell certainly provides a clear formulation of the knowledge intuition—like Russell, Dunne and Meehl he uses the example of a blind person—but it is not clear that he infers from this to the falsity of physicalism.
quoted at the outset.\textsuperscript{14} The central fact about the Mary example is that reflection on it supports two quite separate claims. First, there is the claim that, before her release, Mary knows everything there is to know about the physical nature of the world. Second, there is the claim that, on coming out to her room, Mary learns something. If you grant these two claims you cannot but agree with the knowledge intuition. It is simply part of the notion of learning that if \(S\) learns \(p\) at \(t\), then there is a previous time \(t^*\) such that \(S\) did not know \(p\) at \(t^*\). Thus if you agree both that Mary knows everything physical in her room and that she learns something when she comes out, you cannot but agree that she did not know everything at all in her room. So we might say that the polemical importance of the Mary example derives from the fact that it factors the knowledge intuition into two quite distinct claims—the complete physical knowledge claim, and the learning claim—and that acceptance of both of these claims amounts to acceptance of the knowledge intuition.

3.2. The Thought Experiment Contrasted.
It is instructive in this regard to compare the example of Mary with other examples which also support the knowledge intuition but which are less successful from a polemical point of view.

3.2.1. Broad’s Archangel. Broad says that it is obvious that the archangel did not know something about the physical world. But suppose a mechanist responds by challenging or denying the intuition. Broad would then be in the weak position of insisting that his intuitions go one way, while the mechanist’s intuitions go the other way, and matters would quickly degenerate into a stalemate. What the Mary case does however is break the stalemate in favor of a proponent of the argument. Suppose a physicalist suggests that Mary does know what the smell of roses is like. Then they would be faced with the problem that Mary apparently learns this on coming out. And, if she learns this on coming out, logic alone dictates that she did not know it previously. In a sense, then, Jackson’s version of the argument might properly be called ‘the learning argument’, and it is precisely this that makes it an improvement on Broad’s version.

3.2.2. Feigl’s Martian Super-Scientist. Feigl discusses the possibility that a Martian super-scientist, if he had a “totally different repertory of raw feels”, might not know everything about human beings. As before, however, this example does not improve on the archangel. For we can easily imagine someone insisting that the Martian might be able to figure everything out—indeed Feigl himself insists on just this. And it is unclear how the argument might be developed beyond this rather unsatisfying point. The difference between Mary and the Martian is that Mary learns something phenomenal—in

\begin{footnotesize}
\textsuperscript{14} Thought experiments in philosophy are often based on logically possible but not necessarily realistic scenarios. The Mary case is usually taken as a thought experiment of this sort. However, the case might not be as too unrealistic as one may initially think. For there is a Norwegian vision scientist, Knut Nordby, who is a complete achromat. See Nordby (1996). Nordby visited, with Oliver Sacks, Micronesian island of Pingelap where at least 6% of the population are achromats, and the island of Pohnpei, which has two enclaves populated by Pingelapese, many of whom are achromats. (Compare H. G. Well’s story mentioned earlier.) Their journey is introduced in a book by Oliver Sacks (1996), and a TV documentary film, both of which are entitled The Island of the Colorblind. See also, “The Case of the Colorblind Painter” in Sacks (1985). See also Nordby (1990).
\end{footnotesize}
Feigl’s terminology, she becomes acquainted with a new raw feel—but this is precisely what the Martian does not and cannot do.

3.2.3. Nagel’s Bat. This is an example that has figured prominently in the recent literature—more prominently perhaps even than Mary. Nagel (1974) says that we can never know what the experiences of bats are like because we are too dissimilar to them, and thus we will never know what it is like to be a bat. Something that is at least strongly suggested by Nagel’s discussion is the following consideration. In principle I could know all the physical truths about bats, and yet I could not know about their experiences. So here too we have the knowledge intuition, and the suggestion that the falsity of physicalism follows from the intuition.

However, even if one were to appeal to Nagel’s example to establish a knowledge intuition and argue that physicalism is false, it is still not clear that this is to advance on the simplest knowledge argument. The reason this time is that the bat example relies on the fact we are so different from bats that it is unlikely that we could so much as conceive of or imagine their experiences. But this invites the response that the reason that physical knowledge doesn’t yield phenomenal knowledge in this case is that is is impossible for us to attain the concepts to formulate the relevant phenomenal knowledge in the first place. On the other hand, if this is the explanation for the knowledge intuition, it is difficult see that it places physicalism under threat. After all, physicalism does not entail that we humans must be able to imagine or conceive the experiences of bats.

3.2.4. Robinson’s Deaf Scientist. Mary made her first appearance in 1982. That same year, Howard Robinson produced a version of the knowledge argument founded on the following example:

Imagine that a deaf scientist should become the world’s leading expert on the neurology of hearing. Thus, if we suppose neurology more advanced than the present, we can imagine that he knows everything that 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. (1982, p. 4).

Robinson’s deaf scientist is obviously a close cousin of the Russell’s blind man, though unlike Russell and like Jackson, Robinson is focused on physicalism not empiricism. In

15 Suggested, because, as Jackson (1982) notes, by the end of Nagel’s paper, he renounces any attempt to refute physicalism: “[i]t would be mistake to conclude that physicalism must be false.” (Nagel 1974, p. 446). Despite this, many philosophers (e.g. Lewis (1983), McMullen (1985), Pereboom (1994)) claim that Nagel’s argument is, at its root, identical to Jackson’s Knowledge argument and some even call this style of anti-physicalist argument the ‘Nagel-Jackson Knowledge argument’ (Pereboom (1994)). Jackson himself disagrees: “[i]t is important to distinguish [Nagel’s] argument from the knowledge argument” (p. 132).

16 Of course, one might argue that Mary too lacks a concept while in her room—some philosophers do in fact argue in this way, as we will see. The point at this stage is that it is plausible to suppose that it is psychologically impossible for us to attain the relevant concepts of the experience of bats. Presumably, it is not psychologically impossible for Mary to attain the relevant concepts—on the contrary, she after coming out, there is no question but that she does possess them.
addition, like Mary, the Martian, and the archangel, Robinson’s protagonist is an omniscient scientist who is experientially challenged. (It is interesting to note that Broad’s archangel is also a scientist—Broad’s first version of the argument has physicist Ernest Rutherford in the lead role!) But Robinson’s example is still an example which is not an advance on the archangel. Suppose again that a physicist responded to Robinson by challenging the intuition; we would then reach a stalemate. But with Mary the stalemate is broken.

Of course, one can imagine a development of Robinson’s case in which the deaf scientist is suddenly made to hear again—perhaps this is in fact what Robinson had in mind, though he does not say so. Then one would be able to appeal to the concept of learning in just the way that one can in the Mary case.17 However, while this version of the case is no doubt an improvement, there would then be the independent problem that a person who was deaf and who is now made to hear would change intrinsically in a way that Mary does not change intrinsically when she comes out of her room. Intrinsic changes of this sort are difficult to adjudicate—what happens exactly when someone who was deaf is now made to hear?—and so the Mary example is an advantage even over that case.

3.2.5. Jackson’s Fred. When Mary makes her first appearance, it is almost as an afterthought.18 The main example in the section of the paper that introduces the argument is not Mary but Fred:

Fred has better colour vision than anyone else on record; he makes every discrimination that anyone has ever made, and moreover he makes one that we cannot ever begin to make…Fred can see, really see, at least one more colour than we can: red1 is a different colour from red2…What is the new colour…like? We would dearly like to know but do not; and it seems that no amount of physical information about Fred’s brain and optical system tells us….We have all the physical information. Therefore, knowing all this is not knowing everything about Fred. It follows that Physicalism leaves something out. (1982, pp. 128-129)

Clearly it is possible to develop a knowledge argument on the basis of Fred rather than Mary. What then is the comparison between the two? The most natural thing to say is that Fred is similar to Nagel’s bat—he has an experience which because of various features of our psychological and neural makeup we cannot have. But then we might be

---

17 One sometimes hears the suggestion that Robinson’s case would be an improvement on Jackson’s, because in the Mary case we have the problem that while in her room, she might cut herself, rub her eyes to experience phosphenes (Thompson 1992) or see colorful dreams whereas the deaf will presumably hear no sounds at all. But in response to this, it is not too difficult to imagine away the various possibilities, and in fact, in most presentations of the argument this is taken as read. See, e.g., Chalmers (1996), p. 369, fn. 7.

18 Indeed, it is fair to say that philosophers had largely overlooked the significance of the Mary case until two important critical papers, Horgan 1983 (Chapter 14, this volume) and Churchland 1985a. Jackson’s 1986 paper (Chapter 2, this volume), which solely focuses on the Mary case, was written as a reply to Churchland’s paper.
tempted to say that the reason we cannot know the phenomenal facts about Fred is that it is impossible for us to attain the relevant concepts required to formulate this knowledge. If that is so, however, a knowledge argument which relies on Fred would suffer from the same problem as a knowledge argument which relies on the bat.

While the case of Fred is similar in this respect to the case of the bat, there is still a further respect in which Fred and Mary are similar, while the issues raised by Nagel are different. Unlike Nagel, Jackson is not concerned with the issue of knowing what it is like to be something other than oneself. Jackson contrasts the Nagel example with Fred in the following terms:

> When I complained that all the physical knowledge about Fred was not enough to tell us what his special colour experience was like, I was not complaining that we weren’t finding out what it is like to be Fred. I was complaining that there is something about his experience, a property of it, of which we were left ignorant. And if and when we come to know what this property is we still will not know what it is like to be Fred, but we will know more about him. No amount of knowledge about Fred, be it physical or not, amounts to knowledge “from the inside” concerning Fred. We are not Fred. There is thus a whole set of items of knowledge expressed by forms of words like ‘that it is I myself who is …’ which Fred has and we simply cannot have because we are not him. (1982, p. 132)

This point is connected to one which often comes up in discussions of the knowledge argument. This is the suggestion that the questions about experience that are raised by the argument are at root identical to questions about indexicals and demonstratives that have been widely discussed in philosophy of language. This issue is still very much a live one in the literature. Lewis 1988 (Chapter 5, this volume) refers to the idea as “the third way to miss the point”, but the view is defended at length in Loar (1997, Chapter 10, this volume), McMullen (1985), Bigelow and Paragetter (1990, Chapter 8, this volume) and Perry (2001).

### 3.3. The Structure of the Argument.

So far we have concentrated on one difference between the Jackson’s version and other versions of the argument: the thought-experiment on which his argument is based. The other important differentiating feature of Jackson’s version concerns his defense of the inference, that is, his defense of the step from the knowledge intuition to the falsity of physicalism.

We may approach the issue by noting three points about what Jackson says is a “convenient and accurate way” (1986, p. 293) to present the structure of the argument:

1. Mary (before her release) knows everything physical there is to know about other people.

---

19 It is sometimes suggested that arguments in other areas of philosophy exhibit the same structure as the knowledge argument. For instance, Dlemas Kiernan-Lewis (1991) suggests that an argument against the tenseless account of reality that A. N. Prior’s provides in ‘Thanks Goodness That’s Over’ (1959), is structurally parallel to the knowledge argument.
(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on being released).

(3) Therefore, there are truths about other people (and herself) that escape the physicalist story.

The first point is that we can clearly see here the two claims identified previously as being the ones that reflection on the Mary example prompted: (1) tells us that Mary knows all physical truths; (2) tells us that she does not know all truths because she learns something. So in setting out the argument in this way, Jackson is making explicit the idea that the knowledge intuition is best factored into two.

Second, it is important to notice that the argument—at least in its conclusion—quite explicitly quantifies over truths. Thus the logical form of the argument might be better rendered as:

(4) Every physical truth is such that Mary (before her release) knows that truth.

(5) It is not the case that every truth is such that Mary (before her release) knows that truth.

(6) Therefore, there is at least one truth which is non-physical.

So understood, it is quite clear that the argument is a valid argument. It is as if someone had reasoned as follows: I have examined all the oranges in the box, but I have not examined all the fruit in the box, therefore there must be in the box some pieces of fruit which are not oranges.

However, and this the third point, even if the argument from (4) and (5) to (6) is valid, one might still question whether the conclusion of this argument—and so of the previous argument also—entails the falsity of physicalism, and in that sense one might still question the validity of the knowledge argument against physicalism. For it is reasonable to wonder whether physicalism is consistent with the existence of non-physical truths. In order to illustrate this issue, we need to say a little more about physicalism than we so far have.

3.4 Supervenience and the Psychophysical Conditional
We noted earlier that, in contemporary philosophy, physicalism is usually construed in terms of what is called a supervenience thesis. A simple way to express the idea is as follows. Suppose we gather together all the physical truths of the world into one mega-truth \( P \). And suppose we do the same with all the psychological truths to produce one psychological mega-truth \( Q \). Now consider the conditional formed from \( P \) and \( Q \)—if \( P \) then \( Q \)—and call this conditional the psychophysical conditional. To say that the psychological nature of the world supervenes on its physical nature is to say or imply that the psychophysical conditional is necessarily true. Given the supervenience account of physicalism, therefore, the following conditional is true: if physicalism is true, the psychophysical conditional is necessarily true.

Now, the question of whether the conclusion (6) of the argument above entails the falsity of physicalism turns on the question of whether the necessity of the psychophysical conditional is compatible with there being non-physical truths. On the face of it however, it is difficult to see why these are not compatible. For, as Kripke...
(1980) famously pointed out, being necessary is not co-extensive with being a priori. So it is possible that the psychophysical conditional is a necessary and a posteriori truth. If it is a posteriori, however, it would be reasonable to say that Q is a distinct truth from P, even if they are necessarily connected. But if Q is a distinct truth from P, and P includes all physical truths, then Q is not a physical truth.

In light of the shadow cast by the necessary a posteriori on the validity of the knowledge argument, we should assume that, in both the 1982 and 1986 papers, Jackson was supposing not only that:

if physicalism is true the psychophysical conditional is necessary but also that:

if physicalism is true, the psychophysical conditional is a priori.

And indeed, in subsequent work, this is an assumption which he has spent considerable time explaining and defending (cf. Jackson 1998).20 The key idea is that while Kripke is of course right that there are examples of the necessary a posteriori, it is unreasonable to suppose that the psychophysical conditional is one of them, at least if physicalism is true. According to Jackson, if a statement is necessary and a posteriori, it can be derived a priori from some further contingent statement. In turn, if the psychophysical conditional is a posteriori it too must be derived from some further contingent statement. On the other hand, if physicalism is true, there is no other contingent truth from which the psychophysical conditional can be derived. For physicalism aims at completeness—any contingent truth must already be included in the antecedent of the psychophysical conditional. Conclusion: if physicalism is true, the psychophysical conditional is not only necessary but also a priori.

Whether Jackson is right in this account of the necessary a posteriori is a large issue. Many will question his assumption that any necessary a posteriori statement follows a priori from something contingent. But for our purposes the important points are these: (i) Jackson’s claim that the psychophysical conditional is a priori is not ad hoc but is supported by independent premises; and (ii) if that claim is true, the knowledge argument is a valid argument against physicalism. Putting these two points together, we arrive at the second respect in which Jackson’s knowledge argument is an improvement on earlier versions: his defense of the inference.

4. Responses to the Argument.
We have already taken note of some of the ways in which one might respond to the argument. However, a more detailed assessment of these responses can be arrived at if we adopt the policy, first instituted by Van Gulick (1993, see also Chapter 16, this volume), of organizing the responses around a series of questions.21

Question #1: Does Mary learn anything new?

20 Another response at this point is that it is a mistake to assume that the a posteriori account, even it is can be made out, would be of help in the knowledge argument. For an early discussion of this point, see Jackson 1980.

21 See also Alter (1999).
Everyone agrees that something happens when Mary comes out of her room. Jackson says that what happens is rightly described as learning, and initially this is certainly the natural thing to say. However not everyone agrees on reflection that Mary does learn something.

This issue is the topic of part II of the book. In Dennett (1991; Chapter 3, this volume) we find the suggestion that if we really think about what it is to know everything physical then Mary can know everything before she comes out. Similarly, in his (1998; Chapter 18, this volume) Jackson suggests that the idea that Mary learns something when she comes out is an illusion induced by the fact that the new information is presented as information concerning an intrinsic property, rather than a relational property.

If Mary does not learn anything knew then the argument need proceed no further. The Mary example is no advance on the Broad’s archangel. On the other hand, if we agree that Mary does learn something and thus we answer ‘yes’ to Question 1, we can proceed to other questions.23

**Question #2: Is her learning factual or non-factual?**

We have already noted one sense in which ‘learn’ is closely related to ‘know’—if S learns that p at time t, then there is a time t* prior to t such that it is not the case that S knows that p at t*. But the connections are even closer than this. For example, as is evident from the following, both ‘learn’ and ‘know’ can take both ‘that’-clauses and ‘wh’-clauses in the object position:

(7) Bill knows/learns that the cat is not the mat.
(8) Bill knows/learns how to darn socks.
(9) Bill knows/learns what black holes are.
(10) Bill knows/learns who the President is.
(11) Bill knows/learns where the good clubs are.
(12) Bill knows/learns when the election starts.

According to a tradition instigated by Ryle (1949), there is an important difference between (7) and (8) construed as statements about knowledge. (7) reports a case of genuine factual or propositional knowledge, but (8) reports the presence only of a certain sort of skill or ability—knowledge how versus knowledge that, as Ryle put it. This basic point can be generalized in two ways. First, the point has an obvious application to learning—one might equally distinguish learning how and learning that. Second, knowing how seems naturally to group with knowing what, when and where etc, i.e., with cases in which the object of the verb is a ‘wh’-clause rather than a ‘that’-clause. So we might distinguish more generally between factual learning and knowledge, and non-factual learning and knowledge.

22 See also Jackson (2000).

This distinction opens up the possibility—discussed in Parts III and IV of the book—that the learning that Mary undergoes when she comes out of her room is not factual learning. But if that is so, the knowledge argument fails. For as we saw above, the argument relies on the idea that we can quantify over truths known, and this is precisely to rely on a factual account of learning.

**Question #3: Does she gain only know-how?**

Suppose that what Mary learns is not factual. The question remains as to what sort of non-factual learning it is. The essays in part III discuss the ability hypothesis, according to which what Mary learns is know-how.\(^\text{24}\) According to Lewis (1988; Chapter 5, this volume) for example, while it is true that she learnt something, it is false that there is some new truth about the world—all she learns was how to imagine, identify, and recollect certain experiences.

There are a number of controversial aspects of the ability hypothesis. One is that it is not clear that this is all Mary learns (Jackson 1986; Chapter 2, this volume). A related point (Loar 1997; Chapter 10, this volume) says that since we can embed what Mary learns in the antecedent of a conditional, it cannot be that she learns mere know how. A third objection focuses more directly on whether there is in fact the sort of distinction between knowledge how and knowledge—that that is required by the ability hypothesis.\(^\text{25}\)

**Question 4: Does she merely get acquainted?**

The ability hypothesis is not the only position according to which Mary’s learning is non-factual. Another is the acquaintance hypothesis discussed in Part IV. The acquaintance hypothesis says that what happens when Mary is released from her room is that she becomes acquainted with a new property, the property of being red.\(^\text{26}\) There are a number of different ideas doing business under the label ‘acquaintance’, but the core idea can perhaps be formulated as follows: to be acquainted with the property of being red is to know what red is. The general idea of the acquaintance hypothesis is that one can come to know what red is without knowing any further truths, and that this is precisely what happens to Mary.

---

\(^\text{24}\) The ability hypothesis was originally proposed in Nemirow (1980) as a response to Nagel (1974).


The acquaintance hypothesis is much less discussed than the ability hypothesis, but it is likely that similar sorts of problems will arise for both views. First, it is not clear this is all Mary learns; second, it is not clear that there is a distinction between knowing which and knowing that in the relevant sense. There is here too the additional problem that the notion of acquaintance is an extraordinarily slippery one. Sometimes it is used to denote a certain sort of knowledge as we have noted. But it is also used to denote experience. On this interpretation, however, the acquaintance hypothesis does not move beyond the truism that Mary comes to have a new experience when she comes out of her room.

**Question 5: If the learning is factual, is it learning of a genuinely new fact or of an old fact in disguise?**

Suppose one is persuaded by these criticisms of the acquaintance and ability hypotheses and therefore rejects the idea that the factual/non-factual distinction is the key to the knowledge argument. There is still the question of whether she learns what might be called a genuinely new fact or whether what she learns is an old fact in disguise. The essays in Part V explore the possibility that the latter is the correct way to respond to the argument.27

This sort of position comes in either of two varieties. One version—stated most clearly in Loar (1997; Chapter 10, this volume), Van Gulick (Chapter 16, this volume) and Horgan (Chapter 14, this volume)—appeals directly to the necessary a posteriori. Earlier we pointed out that Jackson is assuming that the psychophysical conditional is a priori, or at least if it is a posteriori then this is irrelevant to the argument. What we find in the papers by Loar, Van Gulick, Horgan are considerations to the effect that he is mistaken in supposing this, and that it is reasonable to view the psychophysical conditional as necessary and a posteriori. Jackson (2002, Chapter 19, this volume), Chalmers (Chapter 12, this volume) and Stoljar (2000) present criticisms of this position.

The other version of the new facts view—presented here in Tye (2000, Chapter 7) but also in Loar (1997, Chapter 10)—focuses not so much on whether the psychophysical conditional is a posteriori, but rather on the idea that Mary gains a new ‘phenomenal’ concept when she comes out of her room. If Mary gains a new concept, then there is a straightforward response to the argument: the reason that she does not know the relevant phenomenal truths is because she does not so much as understand them. In effect, this position makes the charge against Jackson’s version of the knowledge argument that we noted when discussing Nagel’s bat: the objection that the knowledge intuition can be explained psychologically, and if so, there is no threat here to physicalism.

The appeal to phenomenal concepts raises a number of issues of quite independent interest—what are phenomenal concepts, for example, and what it takes to

---

possess them. These are topics explored in Nida-Rumelin (Chapter 11), Hellie (Chapter 13), and Chalmers (Chapter 12). But it also invites a reconsideration of some of the themes already covered. For example: (i) It is sometimes suggested that, at least given the sort of creatures we are, one can only have a phenomenal concept if one has had the relevant sort of perceptual or imaginative experience—this connects with the issue of empiricism and the knowledge intuition noted at the outset. (ii) It is sometimes thought that to possess a phenomenal concept is to have certain sorts of abilities or to be acquainted in the right sort of way with a property—this connects with the ability and acquaintance hypotheses. (iii) It is sometimes said that it is part of the nature of phenomenal concepts that they pick out various properties directly, and not via any mode of presentation—this connects the issue to the question of the necessary and a posteriori.

Question 6. Did she know all the physical facts prior to release?
In questions 2–4 we explored the possibility that Mary does not learn any (genuinely) new facts, either because she did not learn any facts at all, or else because the facts learnt are not genuinely new. But suppose we agree that she does learn some new facts. At this point we might shift focus and consider the other aspect of the Mary example that we have distinguished, the claim that she knows everything physical prior to her release. Of course, according to the usual description of the case, she does know everything. But one might argue that this is a misdescription.

This is a possibility explored in different ways in Part VI. In Horgan (Chapter 14), Stoljar (Chapter 15) and Hellie (Chapter 13) arguments are presented that there is a sense in which Mary might be said to fail to know all the physical facts. If that can be made good, there is no argument here against physicalism, this time for the simple reason that the first premise of the knowledge argument can be denied.28

Question 7. Is Physicalism false?
The final question is simply whether physicalism is false, or rather, whether this is the correct position to adopt in the face of the knowledge argument. This is the position advocated by Jackson in the papers in Part I of the book. In the 1982 paper, for example, the argument is presented as ‘a knowledge argument for qualia’, and in turn qualia are defined as properties of experiences which are non-physical. The upshot is that physicalism is denied and a version of non-physicalism—epiphenomenalism—is true.29

But is dualism really the right conclusion to draw from the knowledge argument? Several philosophers have argued that the argument is so strong that, if successful, it would refute not just physicalism, but anti-physicalism also. For example, Lewis (1988, Chapter 5, this volume) argues that if the knowledge argument defeated physicalism it

---

28 For arguments for the claim that Mary does not know all physical facts prior to her release see Alter (1998), Flanagan (1992), Horgan (1984, Chapter 14, this volume), Nagasawa (forthcoming b), Searle (1992), Stoljar (2001, Chapter 15, this volume), Strawson (1994).

29 Chalmers (1996), Foster (1991) and Robinson (1982, 1993) also subscribe to dualism based on the antiphysicalist conclusion of the knowledge argument, though not all of these philosophers embrace epiphenomenalism.
would defeat substance dualism as well.\textsuperscript{30} Jackson (1986, Chapter 2, this volume) responds that the sort of dualism under discussion is such that no parity of reasons objection is plausible.\textsuperscript{31}

5. Some Further Questions
So far we have concentrated on what the knowledge argument is, and how one might respond to it. However, as we noted at the outset, much of what is interesting here concerns topics the argument raises rather than simply the thesis it attacks. Some of these further topics—the connection to modal epistemology and semantics, empiricism, and phenomenal concepts—have already been canvassed. In this section we will briefly indicate some other topics that are raised by the argument.

5.1. The Scientific Study of Consciousness.
Block (1995) famously distinguishes between two concepts of consciousness: access consciousness and phenomenal consciousness.\textsuperscript{32} He goes on to say that the hard part of the problem of consciousness is the phenomenal notion rather than the access notion. A problem with this suggestion is that the phenomenal aspect of consciousness is notoriously hard to pin down. Here, however, it is natural to appeal to the knowledge argument. Whatever it is Mary does not know about before she comes out of her room, and whatever it is that she learns about when she comes out is phenomenal consciousness, or at least is intimately connected to phenomenal consciousness. To put the point another way, regardless of whether the knowledge argument, and the intuition it is based on ultimately defeat physicalism, they are nevertheless important because they provides a very clear account of what needs to be explained in the study of consciousness.

5.2. The Scope of Scientific Understanding.
While the knowledge argument focuses our attention on what a scientific theory of consciousness needs to explain, it also raises the possibility that explaining it will be beyond us. This theme is raised by Jackson himself in the closing passages of the 1982 paper:

Physicalists typically emphasize that we are a part of nature on their view, which is fair enough. But if we are a part of nature, we are as nature left us after however many years of evolution it is, and each step in that evolutionary progression has been a matter of chance constrained just by the need to preserve or increase survival value. The wonder is that we understand as much as we do, and there is no wonder that there should be matters which fall quite outside our comprehension.

\textsuperscript{30} For a similar claim see Churchland (1985a, 1985b, 1989), Nagasawa (forthcoming a). Churchland notes that he owes “this point to the correspondence with David Lewis” (Churchland 1985b, p. 120).


\textsuperscript{32} Chalmers distinction between the hard and the easy problems is similar. Cf. Chalmers 1995.
Perhaps exactly how epiphenomenal qualia fit into the scheme of things is one such. (1982, p. 135).

Jackson himself mentions this simply as a possibility, as an antidote to “excessive optimism” (1982, p. 136). But some philosophers have gone further, suggesting that the nature of the case is such that we could decide now that the explanation of experience is beyond our ken. This theme has been taken up and defended by McGinn (1989).

5.3. The Definition of Physicalism

The issue of the target of a possible scientific explanation of consciousness, and whether we will be able to carry out the project seem real enough. But a number of writers have questioned whether these issues should be approached via the thesis of physicalism, and perhaps more broadly whether the issue is at root a philosophical issue at all. An example is provided by Noam Chomsky. Chomsky does not discuss the knowledge argument as such, but it is clear that his remarks apply to it:

The mind-body problem can be posed sensibly only insofar as we have a definite conception of body. If we have no such definite and fixed conception, we cannot ask whether some phenomena fall beyond its range. The Cartesians offered a fairly definite conception of body in terms of their contact mechanics, which in many respects reflects commonsense understanding. Therefore they could sensibly formulate the mind-body problem... (1988, p. 142.)

Chomsky goes on to argue that while the concept of body made sense against the particular historical background of the mid 17th century (“the Cartesians”) in later work—particularly the work of Newton and others—those presuppositions were rejected. Chomsky of course does not deny that there might be a commonsensical conception of body or of the physical, nor that the knowledge argument might not be useful in separating out the ordinary conception of the physical from that of the experiential. His position is rather that modern science after Newton has no use for that notion, and instead operates with a conception according to which ‘physical’ in ‘physical world’ has only an emphatic use, much like ‘true’ in ‘true fact’.

Now, in the 1982 paper, Jackson makes it clear where he stands on the question of the definition of the physical: “It is well known that there are problems with giving a precise definition of these notions, and so of the thesis of physicalism. But—unlike some—I take the central problems of definition to cut across the central problems I want to discuss...” (p. 127; see also Jackson 1998). Jackson is surely correct in supposing that to understand the basic intuition behind the knowledge argument we do not need a definition of the notion of the physical—indeed, one might instead view the knowledge intuition as data that any proposed definition would explain. On the other hand, it would surely be strange if the issue of how to understand the concept of the physical had no role to play in the final assessment of the argument. This too, then, is an issue raised by the argument.\(^\text{33}\)

\(^\text{33}\) One defining physicalism see Hempel (1970), Crane and Mellor (1990, 1993), Pettit (1993), Montero (REF), Stoljar (2001a, 2001b, Chapter 15, this volume).
6. Jackson’s Current View.

In 1993 Jackson thought that physicalism was ‘contrary to fact’; by 1996, he had changed his mind: the argument, he said, contained no obvious fallacy, and yet its conclusion—that physicalism is false—must be mistaken. This is one of the more remarkable turnarounds in contemporary philosophy.\(^{34}\) We will close by making some remarks about Jackson’s current view.

In *From Metaphysics to Ethics* (1998), Jackson sets out a vision of how metaphysics should proceed, and then raises some problem cases for it. He begins with the assumption that modern science has within it a certain picture of the world, a picture that is best distilled as the thesis of physicalism. He then notices that physicalism apparently finds no place for a number of items we ordinarily assume to exist—persons, values, freewill, experience. The basic problem that metaphysics sets out to solve is therefore ‘the location problem’—the problem of stating what the location of persons (etc.) is in the world presented to us by science. Now, the most interesting versions of this problem emerge when we are presented with arguments apparently showing that the placement problem cannot be solved—that is, when it is suggested that persons or values as we currently conceive them cannot find a place in the physical world. Thus for example, Moore’s (cf. Jackson 1998 p.150-1) open-question argument reveals something about our concept of value which shows that values-as-Moore-thought-of-them cannot find a place in the natural world. Similarly, Van Inwagen’s (1983; cf. Jackson 1998 p. 44) consequence argument reveals something about the concept of freewill that shows that freewill-as-we-ordinarily-think-of-it cannot find a place in the natural world. In such cases, Jackson says, we have no choice but to settle for something less—to articulate a replacement concept of value or freedom which can take the place of the old. Of course, in both cases, one can imagine also the response that what is needed here is not a replacement notion, but rather a rejection of physicalism, or at least a different approach to the picture of the world implicit in modern science. To this Jackson responds that it is a methodological mistake to suppose that philosophy itself should revise science. That would be to give philosophy and metaphysics an overly “immodest” (1998, p. 42-3) role.

In this context, it is perhaps not too difficult to see the motivation behind Jackson’s change of mind with respect to the knowledge argument. Like the open-question argument and the consequence argument, the interest of the knowledge argument lies in the apparent inconsistency it finds between an item on the troubling list and physicalism. But to adopt dualism in response to this argument would be to let philosophy play an immodest role, and this is something that Jackson rejects in parallel cases. In short, there seems to be a pattern in Jackson’s thinking on these topics, and his earlier position represents an important departure from this pattern.

There are many questions to be discussed here about Jackson’s change of mind and his conception of philosophy, some of are taken up in this book. Is it really true that

---

\(^{34}\) His conversion became big news in the philosophy community, and was reported in the Philosophy News Service with the headline ‘Frank Jackson changes his mind’: “Frank Jackson is now a physicalist. Yes, the same Frank Jackson who used to call himself a ‘qualia freak’ now claims that qualia (the visual sensation of red, the feeling of pain, and other sensory experiences) can be given a complete physical explanation. After more than a decade of defending his anti-physicalist views, Jackson has come around to the other side: physicalism, he now claims, is true.” (Montero. 1999.)
there is a pattern in the questions that Jackson is interested in? Is it really true that the necessary a posteriori is to be analyzed in the way that he says? Is it really true that physicalism is the picture of the world implicit in modern science? What exactly are the commitments of physicalism anyway? What are the replacement notions that Jackson articulates and to what extent can they play the role of the original notions? What does it mean for one notion to play the role of another? And—what is most important for this book—if there is a mistake in the knowledge argument, just where exactly does it lie? All of these fascinating questions—including of course the question of what the scientific explanation of consciousness is—bubble to surface when one discusses the knowledge argument.