

Pyrrho Press
Forthcoming Titles

Science and Religion (1997)

American Philosophers (1997)

Australian Philosophers

Edited by

Phil Dove

Moira Nicholls

Leila Shotton

Pyrrho Press Hobart
1996

here now, he would have much to say in defence of his theory?
I can only leave you to speculate on whether or not he would
convince you.

4 Casting out Demons and Exorcising Zombies: Exposing Neocartesian Myths in Frank Jackson's Philosophy of Mind¹²

Jay Garfield

For we could think, feel, will, and remember, and we could
also "act" in every sense of that word, and yet none of this
would have to "enter our consciousness" (as one says
metaphorically). The whole of life would be possible
without, as it were, seeing itself in a mirror. (Nietzsche,
The Gay Science, 354)

Introduction

Australian philosophy of mind is best known for the so-called "Australian Materialism" of J.J.C. Smart (1959) and D.M. Armstrong (1968, 1980). Frank Jackson, who taught for many years at Monash University and is now on the Research Faculty of Social Science at the Australian National University, is best known for his attack on the central tenets of that school, and for his revival of non-physicalist theories of mind. In fact, Jackson is justly famous in contemporary philosophy for his espousal of unpopular theories, including his view that (at least some) mental facts and properties are non-physical; his commitment to a sense-data theory of perception; and his commitment to anti-individualistic theories of mind (Jackson & Pettit, 1988).³ The former two represent revivals of positions long thought to be dead and buried in the philosophy of mind; the last has been making steady progress into the mainstream, but is still regarded by many with great suspicion.

In this essay I will focus on Jackson's anti-physicalist view of mind and on his sense-data theory. These are the two doctrines for which he is best known, and his work in these areas is certainly his most influential. I also choose to focus on these two theoretical commitments because they are deeply connected to each other, and reflect a vision of mind, of knowledge, and of self-knowledge that is deeply embedded in our collective self-image—a vision I find to be both immensely plausible and deeply false. According to that vision, we have only *mediated* access to external phenomena through special cognitive intermediaries—representations, or sense data. We have access to ourselves and to these representations or sense data, on the other hand, *immediately*. It is not obvious that this view of mind is implicated in these two strands of Jackson's thought, nor that they are linked thereby. I will therefore first explain and criticise each view independently, and only then turn to this more global assessment of Jackson's philosophy of mind.

11 pursuing this philosophical investigation I must emphasise that though my aim is critical, I have the greatest respect for Jackson's work, and consider it to be of the first importance to contemporary philosophy of mind. Cresswell (1980) notes in a perfectly serious tone that among the great virtues of Frank Jackson's writing is that he articulates a position with sufficient clarity that one can see exactly what is wrong with it. In philosophy this is a great virtue indeed. For often—especially in cases where the position in question is intuitively plausible—philosophers cover subtle and important errors in imprecise or obscure formulation. This impedes philosophical progress, which always depends upon the dialectical engagement of opposing positions with one another in a conversation whose goal is, *inter alia*, increasing clarity regarding the concepts under discussion. So, whatever one's position in the debates with which Jackson engages, one is indebted to him for raising the level of discussion in a way characteristic of the very best philosophical thinking and writing.

1. *The Knowledge Argument*

The "knowledge argument" is among the most famous single arguments in the philosophy of mind literature of the last quarter-century. It is first articulated in Jackson (1982) and refined in response to a variety of criticism (Churchland, 1985; Horgan, 1984; Levin, 1986; Lewis, 1983)⁴ in Jackson (1986). Here is the original formulation:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room *via* a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and uses terms like 'red', 'blue' and so on. She discovers, for example, just which wave-length

how this produces *via* the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue.' ...

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

The argument, when flensed, runs as follows:

1. When Mary was still confined to her black and white room, she knew everything there was to know about the physical facts regarding perception.

2. When Mary left the room and saw her first tomato, she learned something new, *viz*, *what it is like to see red*.

So,

3. *What it is like to see red* is not a physical fact about perception.

So,

4. Qualia (the felt character of perceptual states) are non-physical, and the complete physical story about reality is not the complete story.

This argument is interesting in a number of respects. Let us enumerate seven possible replies, each of them a plausible refutation, before getting to the heart of the problem: One feature that has attracted considerable notice (Churchland,

conclusion about the nature of things—from epistemological premises—that is premises regarding the nature of our knowledge. That one should be able to get from such premises to such a conclusion is at least surprising. A second difficulty with the argument is that it does not rely on any precise specification of the difference between the physical and the non-physical of the kind one might expect to find in an argument to the effect that certain things are definitely non-physical.

Third, one might be suspicious of this argument because it at least appears to commit what philosophers call "the intentional fallacy." That is, it trades on descriptions that occur within the context of verbs like "knows." These verbs create contexts in which we use descriptions without being committed to the possibility of using them outside of those contexts⁵. So, consider the following argument:

1. John knew that Queen Elizabeth II just walked by.

2. Queen Elizabeth II is the richest woman in the world.

So,

3. John knew that the richest woman in the world just walked by.

Now, this argument is obviously invalid. John may have been under the common misapprehension that Jackie Onassis was the richest woman in the world. The problem with this argument is that it uses the expression "Queen Elizabeth II" both inside and outside the intentional context created by "knew." The knowledge argument also involves intentional contexts, and hence generates suspicions of invalidity on those grounds.

A fourth reply runs as follows: The knowledge argument assumes that in the black and white room Mary could really know all of the physical facts about perception *and still not know what it is like to see red*. But this just betrays a failure of our

current context. Now, to be sure, given what we currently know about perception it is hard to imagine our studying physiology textbooks and somehow coming to know what it is like to see red, or what a pineapple tastes like. But, one might say, this merely reflects the poverty of our current knowledge of neuroscience, and our inability to imagine a more highly developed neuroscience of the kind known by Mary. Mary, after all, knows *all* of the physical facts about the physiology of perception. So, it could well be that *in virtue of knowing them* she knows what it is like to see red. Why? Well, perhaps because it is one of those physical facts. Maybe not, but to assume that is not for the purposes of the argument is to assume the very conclusion the argument is meant to establish. So to even state the example in a way that entails the conclusion is to beg the question (Churchland, 1985; Conee, 1994).

Fifth, as a kind of mirror image to the last objection, one might concede that Mary could not, no matter how fine an education she receives in the black and white room, come to know what it is like to see red. Perhaps, one *does* need to see red in order to know what it is like to do so, just as our naive intuitions predict. But then why not say that it follows that—since *what it is like to see red* is a physical fact (or at least could be one)—Mary just *can't* learn all of the physical facts about perception in the black and white room (or at least perhaps can't)? To assume that she can—despite the impossibility of her acquiring qualitative knowledge—is just to beg the question at issue. That is, it is to assume at the outset what the argument purports to prove, *viz.*, that qualitative facts are non-physical.

Sixth, it has been suggested that Mary, upon exiting her academic prison, gains no new knowledge at all. She gains a new experience, all right: the experience of seeing red. But *what it is like to see red* is not an object of knowledge at all. It is only our untutored, rough and ready, folksy description of a brain state. And Mary, *ex hypothesis*, already knew all there

was to know about that brain state, so the previous argument that asserts that Mary learned something new when she first saw red is simply false (Lewis, 1983).

Finally a related objection runs as follows: Indeed, Mary has not gained new knowledge when she leaves the room. She already knew all there was to know about the neurophysiology of perception, *including what it feels like to see red*. But she knew it under a different description— a physiological description (just as our erstwhile friend John knew the Queen under her royal, but not under her financial description). But she does gain *something* new when she leaves the room, namely a set of *abilities*. These include the ability to describe the experience of seeing red from the first-person point of view, the ability to recognise red objects, etc. But none of these new abilities constitutes, according to proponents of this objection, new knowledge (Churchland, 1985; Dennett, 1991; Levin, 1986; Lewis, 1983).

Now, I think that there is something to be said for each of these objections. Perhaps Jackson can mount satisfactory replies to some of them. I doubt that all of them could be dismissed, and a few cut right to the centre of the argument. But I want to leave them aside for now and to concentrate on a different problem— one that I think reveals most clearly the fundamental error that motivate Jackson's and related views, and that reveals most starkly why it is an error.

2. The Zombie Problem

I will now develop the central argument of this essay, and what I think is the most decisive refutation of the knowledge argument. In order to articulate this argument, I must introduce the notion of a zombie. The original idea and the term derive from Haitian voodoo. But I make no claims to the authenticity of my version. A zombie as I will understand it is a human being who is just like you or me, but who lacks all qualitative states.⁶ To put it simply: While there is something that it is like to be you, or me, or Frank Jackson, or perhaps

other respects. They act like us, talk like us, and share our beliefs. Indeed, by stipulation my zombie twin would say *everything* that I say, and would *believe* everything that I believe, including (reality to the contrary notwithstanding) that it (like me) has a rich inner life. But while I am correct in this last belief and assertion, my zombie twin is wrong.⁷

With this sketch of zombie psychology in hand, we can sketch the zombie problem for the knowledge argument:

1. If Jackson is right, the complete physical story about us leaves out qualia.
- (That, after all, is the conclusion of the knowledge argument.)
2. If that is true—that is, if the complete physical story leaves something out, it follows that there could be (in the metaphysical sense of “could be”—that is, it is logically possible that there is) a world exactly like ours in all physical respects, but in which there are no qualia.

Call this world “zombieland,” since in that world, everybody is a zombie, even though to the untutored eye zombieland and Earth would be completely indistinguishable. Note that the possibility of zombieland is simply a way of restating Jackson’s conclusion—that somebody could be physically just like me, but could differ from me in the respects that the physical story leaves out. To differ in those respects that the just be to lack qualia—to be a zombie.

3. These zombies believe and sincerely assert that they have qualia, even though they don’t have them. They have no inner life. But despite that fact, since they are behaviourally just like us, and have the same non-qualitative states that we do (including beliefs) they—like us—believe and assert that they have qualitative

these assertions, they are wrong. Undead wrong.

4. This raises the crucial question: **How do we know that we are not zombies?**

5. Jackson must be able to answer this argument if the knowledge argument is to be successful at all. For if Mary (or we) are to know what it is like to see red or to taste chocolate, we must know that we have seen red or tasted chocolate and have had the relevant qualia.

But

6. We can only know that if we know that we are not zombies.

(For if we are zombies, then we have never had these experiences, and hence cannot *know* what it is like to have had them.)

7. We cannot know that we are not zombies simply in virtue of our *belief* that we have had qualia. For zombies share that belief. Nor can we simply say that when we introspect we find those qualia. For zombies sincerely believe that when they introspect they, too, find qualia.⁸

So,

8. If zombies are possible, we can’t know that we aren’t zombies!

(The mere possibility of the existence of zombies would entail that we can’t know whether or not we are zombies, since all of our beliefs, including the belief that we are not zombies, would be shared by the zombies themselves (ourselves?).)

So,

impossible. This is logically equivalent to (8).

And

10. The knowledge argument only succeeds if Mary in fact learns something new when she emerges from the black and white room, and hence only if Mary knows that she is not a zombie.

(Note that if Mary were a zombie, she would gain no new knowledge. So if she does gain new knowledge, she must not be a zombie. And if that is to be knowledge, she must *know* that she is no zombie.)

But

11. If we and Mary know that we are not zombies, and hence that zombies are impossible (otherwise we couldn't know that we are not zombies) then the complete physical story leaves nothing out.

This follows directly from (2).

So,

12. If the knowledge argument shows *anything* then the complete physical story about the world leaves nothing out.

This is so because the knowledge argument only shows anything if Mary gains the requisite knowledge upon leaving the black and white room; but she only gains that knowledge if she also knows that she is not a zombie; she only knows that if zombies are impossible; and they are impossible if, and only if, the complete physical story leaves nothing out.

So,

13. If Mary learns anything, then what it is like to see red, and qualia in general, are physical facts.

This is a long argument. If it is correct, it shows that the knowledge argument cannot succeed. (It also explains why many of the seven objections canvassed earlier also succeed, but that is a long story, and beyond the scope of this essay.)⁹ The main point is this: Either qualia are physical or they are not. If they are physical, the conclusion of the knowledge argument is wrong. But if they are not physical, we could never come to know anything about them, and hence the knowledge argument cannot establish either their existence or their status. Or to put this point another way: Either Mary learns something when she leaves the room or she doesn't. If she does, the facts she learns are physical. If she doesn't, there is no new knowledge to make the argument go.¹⁰

3. What Does all this Zombie Talk Gain Us?

The knowledge argument is important, persuasive and revelatory precisely because it suggests both that there is more to us than our physical being, and that an examination of the nature of our knowledge demonstrates that. There are three important theses here, and each deserves a bit of emphasis. First, there is the obvious conclusion: The complete physical story is not the complete story. There are non-physical facts about persons. This is a dramatic, and not a trivial claim. Now it is not tantamount to the claim that we have immaterial souls, a claim with which it might easily be confused. That is, it is not a *substantialist* claim in any sense. Nonetheless, it is an ontologically significant claim. It is a claim that there exist non-physical facts in addition to all of the physical facts. What the nature of those facts are, and how exactly they are related to physical facts remains unstated - but that at least becomes a matter for metaphysical, and perhaps scientific, investigation.

Second, the argument suggests that we can establish the reality of this domain of non-physical facts by examining not that reality or those facts directly, but rather through reflection purely on the structure of our knowledge. That is also a

unological conclusions through epistemological arguments, and this because of the ubiquity of intentional fallacies in this terrain. There may indeed be exceptions, and this might be a case in which this is possible. But as we have seen, there is at least *prima facie* reason to believe that precisely such a fallacy is committed here. Whether or not that charge can be made to stick, we can see that it is a substantive claim implicit in the knowledge argument that in this case such epistemological reflection is sufficient to establish a very strong ontological conclusion.¹¹

Third, and perhaps less obviously, this argument relies on alleged special features of self-knowledge. It would be impossible for Mary, according to the knowledge argument, to come to know what it is like to see red in virtue of someone else's coming to see red, and her learning about that person's experience. The alleged object of this new knowledge—the non-physical fact of which Mary becomes newly aware—is hence a *private* fact, a fact available only to first-person introspective knowledge.¹²

Now this third, least obvious feature is in fact the most important, and is the one that the zombie argument brings from the conceptual sediment in which this argument is grounded into the light of day. For the putative difference between zombies and us is just this: We have these private objects and they don't. But the Zombie Argument, through demonstrating the impossibility of zombies and the consequent completeness of the physical story shows that there can't be any such things that we can have but zombies can't. And so, despite the overwhelming plausibility of the view that we have these private, inner objects of knowledge which themselves are non-physical, but which somehow mediate our knowledge of the physical world, the zombie argument shows that there are no such objects, and that knowledge can have no such structure.¹³

This view about the nature of knowledge is very old and very deeply ingrained. It is introduced most explicitly and

articulately by Descartes in the famous piece or wax argument in the second *Meditation*. It is deeply bound up with what Sellars has called "The Myth of the Given," the idea that our knowledge is grounded on immediately given introspective knowledge from which we then extrapolate more dubious, mediated empirical knowledge of the external world. (Sellars, 1963a, §§45-50). The knowledge argument is *inter alia* an expression of this Cartesian form of the Myth of the Given, and our refutation of it is *inter alia* our liberation from the tyranny of that myth.¹⁴

For the Zombie argument shows that if there is anything about us that transcends the physical, we could never know it. This strikes at all three of the features of the knowledge argument just noted and does so in virtue of striking to the core of the Cartesian view of the immediacy of self-knowledge as contrasted with the mediate character of our knowledge of external phenomena. The zombie argument shows that there are no such immediate non-physical objects of inner awareness knowledge of which Mary is supposed to gain. It also shows that any reflection on our knowledge can only lead us to objects of knowledge that are in principle public in a special sense—objects that are in principle, even if not in fact, knowable to another.

We must also emphasise what the zombie argument does *not* show. It does *not* show that there are no qualia. In fact it assumes that there *are* qualia—that at least some mental states do have felt qualities. That is what it is to say that we are not zombies. Moreover, it does *not* purport to show that Mary had or even could have had red qualia in the black and white room, no matter how much she knew about the perception of red. It may well be—and this would be an empirical, not a conceptual matter—that the only way to have red qualia is actually to see red. So, since Mary saw no red in the black and white room, it would perhaps be impossible for her to have had red qualia in that room.¹⁵

More importantly, the zombie argument does *not* show that Mary learned nothing new when she saw red. Before

sne saw red, she did. So of course she gained new knowledge when she left the room. About that, I agree with Jackson (and so, importantly, disagree with certain of the critics noted above, especially Lewis (1983)). In fact, anyone who reflects a bit on the zombie argument should see that it *presupposes* exactly that there are qualia, that Mary didn't have them in the black and white room, and that she did learn something new when she came out. Otherwise, she couldn't know that she is not a zombie.

So what *does* the zombie argument show, and how does its conclusion differ so dramatically from that of the knowledge argument? Where, that is, do I differ so sharply from Jackson, other than on the soundness of particular philosophical arguments? Precisely in this: If we accept the zombie argument, and the morals of that argument drawn above, we must accept that while Mary learns something new- above, we acquire new knowledge- there is no new *object of knowledge* revealed, no new (non-physical) facts discovered. Rather, Mary gains a new *perspective* on the same facts she knew about before. Whereas once she knew those facts- the facts about the perception of red- from a third-person perspective, after leaving the room she comes to have a first-person experiential perspective on those same facts. We might compare this to the case of our friend who sees the Queen go by and later comes to see her bank statement. There are not two Elizabeths about which he now knows- one regal and the other wealthy, but rather one- with respect to whom he has two very different perspectives. A change in perspective does constitute new knowledge- but not necessarily of new facts or objects; possibly only of our cognitive relation to the same old phenomena.¹⁶

In characterising the change in Mary's epistemic state this way, I am refusing to draw the metaphysical conclusions from her change in state that Jackson invites us to draw. So another moral of this argument is the re-emphasis of a warning we sounded at the outset: Beware of metaphysical conclusions

drawn from epistemological premises. We can describe the metaphysical facts in a perfectly ordinary way, consistent with the exotic epistemology, but without introducing any non-physical facts or entities.

4. *A Very Brief History of Sense Data Theory*

It would be nice to be able to begin this discussion with a neutral definition of sense-data, commonly accepted as at least a starting point by all parties to the sense-data debates, past and present. Alas, that is impossible. For not only do proponents of sense-data theories disagree with their opponents regarding whether there are indeed sense-data, but proponents often disagree among themselves about what sense-data are, what they do psychologically and theoretically and why we should posit them. On the other hand, I don't want to digress into a lengthy discussion of the historical details of sense-data debates and adjudicate between rival theories of their nature. So I will simply present what I take to be the most influential, classical version of sense-data, in part because of its status as dominant in the most important debates about their existence, and in part because attention to the classical version can highlight several central features of Jackson's account- some because they are shared, and some because they contrast.¹⁷

It is important to begin by noting that in this case—just as in the case of non-physical qualitative facts—we find Jackson swimming against a powerful prevailing current in contemporary philosophy of mind. Sense data theories had their heyday in the middle of this century, and by the mid-1960's were regarded as historical curiosities. Jackson's revival of a sense data theory in (1977) hence attracted a great deal of notice in part because of its clarity and acuity, but in part because of its startling anachronism.

Roughly speaking, sense data in the classical formulation of the theories are the immediate input to sensory consciousness. They are meant to be the first items registered

Now, even this rough, preliminary sketch raises a host of problems: Are these causal and epistemic demands on sense-data consistent with one another? Are sense data physical, psychological, or neither (or both)? Are physical objects known via sense-data or constructed out of sense data? Are we aware of our sense-data in introspection, or are they theoretical entities? And others besides. But let us leave at least most of these aside for a moment, and attend to the reasons that led philosophers to posit sense data in the first place. We will then turn to the considerations that led to the demise of classical sense data theory.

The first motivation for sense data theory is the epistemological position called "foundationalism." This is the plausible idea that empirical knowledge must have a foundation. If, the foundationalist argues, every item of empirical knowledge rests for its justification upon another, then the whole edifice of our knowledge of the world is unsupported, and unjustified. In such a case *s/he* argues, we really know nothing at all, since to know something is, *inter alia*, to be justified in believing it. Now, if I am to know such a fact as *that there is a brown cat on the table* it seems that the appropriate kind of foundation is my immediate knowledge of the existence of the relevant sense data. I can then infer

these sense data to the existence of the cat (either by literally constructing the cat out of sense-data or by inferring the existence of the cat from the sense data as evidence, depending upon the version of the theory I prefer). Since our empirical knowledge rests upon our perception of material objects and their properties, and since we don't directly perceive these material objects and properties, the sense data theorist argues, that knowledge can't serve as an ultimate epistemic foundation. But sense data are known directly, not inferentially, and can serve as the basis of our knowledge of material objects and their properties. Since we do have such knowledge, it must have a foundation, and so there must be sense data. (Note another epistemological argument for an ontological conclusion.)

A second classical reason for positing sense data goes back to a model of perception due to Kant- that of perception as a synthesis of percepts on the basis of given raw material. In perceiving a material object, such as a cat, Kant and his followers argue, we don't find a cat literally given to our mind, somehow pushed through our eyeballs into our brain. Rather, some basic sensory information is given to our perceptual system, transmitted to our brain or mind (depending upon one's psychology and ontology) and there synthesised into the percept we recognise as a cat. But you can't synthesise a whole cat out of nothing. Synthesis presupposes raw material as input to the sensory systems. That raw material, the sense data theorist argues, is a manifold of sense-data.

A final important motivation for sense data theory (there were more, but this will be sufficient for present purposes) was the so-called "argument from illusion." Consider several pairs of perceptual situations: In one case I see an actual brown cat. In another I see an illusory brown cat as a consequence of optical illusion accomplished with holograms, mirrors or hypnosis. In another case I see a genuinely bent stick and then a straight stick made to look bent in virtue of being dipped in the water. Finally, imagine seeing a round coin seen on edge

of these pairs of cases is a pair of situations each of which is identical to the other in respect of sensory experience. But the material objects that are the *prima facie* objects of perception are different in each situation, for each pair. If we want to account for the sameness in the cases then, we need to posit an object of awareness distinct from the material objects in question that is invariant despite the variations in distal objects—a common core to veridical and illusory perception. That object of awareness—the common core—the sense data theorist says, is the set of sense data. So the sense data caused by a real cat and a hypnotically presented cat are the same; by a bent stick and a dipped stick the same; by a round coin on edge and an oval coin the same. That is why these things look the same—we perceive them not directly, and hence not with a direct awareness of their differences, but indirectly, through sense data that are their common effects.

Given these powerful reasons for believing that there must be sense data as the psychological foundation of perception and as the epistemic foundations of perceptual knowledge, why did they fall out of favour? There are at least four distinct principal reasons, and perhaps a host of subsidiary considerations. Again, the arguments are too complex to do them complete justice here. But we can give their general flavour. The first reason is that foundationalism itself fell out of favour. Arguments due to such philosophers as (Quine, 1951; 1960; Sellars, 1963a; 1963b) emphasise the seamlessness and mutually reinforcing character of knowledge, and suggested that there is no such thing as immediate knowledge that requires no justification. Even knowledge regarding sense data, these philosopher argued, rests on concepts of colours, etc that could only be acquired through a wide range of broader empirical knowledge (Sellars, 1963a, §§10-23).

A second reason for the demise of sense data theory was that philosophers began to have doubts about whether sense data were possible objects of knowledge at all. After all, we have some grip on what it is to know about physical things,

and what the structure of our knowledge is, and that a cat is on the table. And we know how to acquire that knowledge. But it is far from clear just what the structure of our knowledge of our sense data is, or how that knowledge is acquired.

A third, and closely related problem raised for classical sense data theories concerns their inaccessibility to introspection. After all, one of the principal motivations for introducing sense data was their supposed immediacy as objects of knowledge, as opposed to the allegedly mediated character of our knowledge of ordinary physical objects. So it should turn out that when we introspect—drawing our attention to the immediate contents of consciousness as opposed to the external objects of perception—we should find these sense data present to the mind. But we don't. At least it is not clear that we do, and if the picture of the structure of our knowledge painted by the classical theory is anywhere near correct, we should.

Finally, ontological worries about the status and nature of sense data raise serious difficulties for any sense data theory. It is just not clear *what* sense data *are*, or *where* they are to be found. Are they properties of physical objects, to be found in space, or on the surfaces of things? Some theorists (Moore, 1953) argued that they are. But if so, it is not clear how they can serve their principal functions as immediate objects of perception and knowledge.

Are they then psychological objects to be found in the mind? That would give us the immediacy we want, but it would sever their connection to the objects about which they allegedly give us mediated knowledge. This would be particularly acute for those who would argue that physical objects are literally constructed out of sense data (Ayer, 1940), but would also be a problem for a more moderate sense data theorist for whom sense data merely function as evidence about physical objects (Broad, 1925; Price, 1932). For it seems that if they are purely mental, we would have no reason to

however, making them purely mental makes it hard to figure out how they could have all of the perceptual properties they are meant to have: Are thoughts red? Oval? On the table?

Or are sense data neither physical nor mental? Well, there are some things plausibly like that: abstract objects such as freedom, or the number seventeen. But sense data are not abstract entities like those. If they were, they could not perform any of their metaphysical or epistemological functions. Is there another sense of being neither mental nor physical? None seems to be readily available.

For these and related reasons, sense data theory fell out of favour, and was replaced as the dominant view in epistemology by various naturalistic non-foundationalist approaches utilising either direct theories of perception or constructivist representative theories. It is against this background that we must understand Jackson's own version of sense data theory. As we shall see, it incorporates some features of the classical view, but is distinct from it in important respects as well.

5. *Jackson's Revival of Sense Data*¹⁹

Jackson devotes an entire closely argued book to the exposition and defense of sense data theory. It would be folly to claim to a complete exposition here. Rather, I will develop what I (following (Cresswell, 1980)) regard as Jackson's three central arguments for the existence of sense data and for their centrality to the theory of perception, and then explain the most salient features of sense data themselves as Jackson characterises them.

The first argument is the "Looks Argument." (Jackson, 1977, pp. 30-49, 109-111). Suppose that Frank is looking at a lovely ripe tomato in broad daylight. Macbeth, on the other hand, is in his castle late at night in Act II, famously hallucinating a bloody dagger ("Is this a dagger I see before

me, its blade bent double in my hand? Come, let me clutch thee...") Now, consider the following pair of sentences:

- i) That tomato looks red to Frank.
- ii) That dagger looks red to Macbeth.

1. The "looks" in these two sentences cannot be comparative. That is, they can't be analysed as "looks the way red things normally look to Frank/Macbeth."

(This could be so for a number of reasons. Suppose for instance that Frank is colourblind, and most red things look green to him. But this tomato is different, and for some reason, in this light, looks red. Macbeth may never look at daggers, or at red things. There may be no way they normally look to him.)

So,

2. "Looks" in these two cases appears to denote a basic relation obtaining between Frank and something seen and Macbeth and something else seen. And it appears to denote the same relation in both cases.

But,

3. That thing can't be a physical object, since in Macbeth's case there is no dagger, and we said that the relation is the same in both cases.

So,

4. The thing to which they are each related that is responsible for the common core of their sensory experiences is a set of red sense data.

Note that this argument is closely related to the argument from illusion. But it is slightly different in its emphasis on our understanding of the word "looks." It also supposes that sense data are functioning as immediate objects of perception, but

to the second of Jackson's principal arguments, the "Many Properties Argument." (Jackson, 1977, pp.60-74) Consider the following pair of sentences:

- iii) That tomato looks red and round to Frank.
- iv) That dagger looks red and sharp to Macbeth.

- 1. The pairs of properties denoted in each of these two sentences must be properties of the same things. That is, the same thing must be both red and round in Franks's perceptual experience, and the same thing must be both red and sharp in Macbeth's.

But,

- 2. These properties can't be in the objects, for again, Macbeth's perceptual experience relates him to no external object.

And,

- 3. They can't be properties of sensations (or anything else mental). Because even if we thought that sensations (or other mental states) could be red, nobody would assert that there are round or sharp sensations or states of mind.

So,

- 4. These must be properties of sense data. Since sense data are the only plausible perceptual object that could serve as the common locus of all of the properties of these percepts.

Note that in this argument Jackson is explicitly assigning sense data a non-physical, non-mental ontological status. This curious neither-fish-nor-fowl understanding of their mode of existence is unique to Jackson's formulation, and is his way of

theory by its horns. We now turn to the final of Jackson's three principal arguments, the "Colourlessness of the World" argument (Jackson, 1977, pp. 122-137). This argument has its roots in independent commitments in the philosophy of science, commitments we will simply note without comment, since to enter that terrain would involve a long digression.²⁰

- 1. We believe that tomatoes, etc are coloured because of our experiences of them.
- 2. The immediate causes of these experiences are events in our brains.

(That is, our experiences, whatever their distal causes may be [including tomatoes or actual daggers] are caused via a chain of events whose physical terminus is in our brain. Another way of making this point is that, in any particular case, even if the distal causal events had not happened, but the brain events had somehow occurred, we would have had the same phenomenal experiences.)

- 3. These brain events are the effects of the causal properties of the distal objects of perception.

(After all, since they are effects, they must have causes, and any property of an object that could be a cause must be a causal property.)

- 4. An object's causal properties are those that the best science of the domain in which that object occurs tell us it has.

(That is what science is for- to tell us about what causes what, and what the causal properties of objects are.)

But,

- 5. Redness, roundness, sharpness, etc are not properties recognised or ascribed by any science.

charge, ion concentration, dendrite branching, etc. Tomatoes, daggers, etc are physical objects, and hence have only the causal properties recognised by physics.)

So,

6. Redness, roundness, sharpness, etc are not responsible for our experiences.

(This follows from (3), (4) and (5).

So,

7. Our experience of objects gives us no reason to ascribe these properties to them.

And

8. They are not in the objects.

(From the previous two arguments, now buttressed by noting that our experiencing of the objects gives us no reason to locate them there.)

And

9. They are not properties of sensations
(from the previous two arguments).

So

10. They are properties of sense data.

Now this argument has a kind of surprising ring to it: We are asked not only to believe in sense data, but also to believe that none of the things we naively take ourselves to have seen all these years actually have the properties we have naively ascribed to them. The tomato is not in fact red; the cat is not in fact brown; the stick is not in fact straight. It is interesting to

properties (the tendency to take science not only as the ultimate but as the only arbiter of the real) is accepted not only by Jackson, but by many of his critics, such as Churchland (1985) and used as an argument *against* the reality of sense data, as *they* are not posited in any of these sciences either! But surely, if neither the objects themselves nor our sensations are possible loci for the properties we perceive, sense data must at least be taken as plausible loci. We will return to criticism of this argument, along with its two companions, below.

Let us close this exposition of Jackson's account by noting three distinctive features of sense data on Jackson's view-features that mark differences between this and more classical versions of the theory. For one thing, Jackson's sense data are not mental objects in any sense. They are not sensations, or percepts in the mind. Nor are they physical objects. Rather they are non-physical constituents of the world. In this way they can be compared to the kinds of non-physical facts that Mary allegedly came to know. Second, Jackson's sense data are not immediate objects of knowledge, or at least are not privileged as such as part of a foundationalist epistemological programme. Rather, we come to know of these sense data only through sophisticated philosophical reflection. Finally, sense data for Jackson are not introspectible. He nowhere claims that these things can simply be found in inner sense- they are not given to us when we look inside. Hence the objection often raised against classical theories that sense data ought to be, but are not, introspectible has no bite against Jackson's theory which posits them more as theoretical entities.

6. *Why Jackson's Arguments for the Sense Data Theory Fail*

Sophisticated and compelling though these arguments and the theory they underpin may be, neither do the arguments succeed nor can the theory be maintained. Let us look first at

errors embodied in the theory itself.

We begin with the "Looks Argument." It can be seen as a defense of the following inference:

(L1) There are red sense data because the tomato and the dagger look red.

Now, this inference is not, to be sure, presupposed in the argument. That would be to beg the questions rather baldly. But to the extent that the argument succeeds, this must be a good inference. We can state this inferential principle more generally as follows:

(L2) If an object looks to have property P, there are sense data that do have P.

If this argument is valid, (L2) must be true. Note, however, that the tomato also looks tomatoish, and the dagger also looks daggerish. It would follow from the validity of the central inference from things looking to have a property to the existence of sense data that *do* have that property that the following argument is sound:

1. The tomato looks to be a tomato; the dagger looks to be a dagger.
2. If an object looks to have property P, there are sense data that do have P. (L2)

So,

3. There are sense data that are tomatoes and daggers.

Of course this conclusion is absurd, and would contravene central tenets of Jackson's theory of sense data, such as their non-physicality. Moreover, if sense data are simply physical objects renamed, there is no reason to go through this long theoretical song and dance: just say that we perceive physical objects and be done with it. But the whole point of this theory is to interpose something between us and the objects we perceive to solve the of the problems about perception noted

above. If sense data just turn out to be those objects, nothing is accomplished. The theory hence looks both pointless and internally inconsistent (Cresswell, 1980, p. 123-31).

Consider now the Many-Properties Argument. It also can be seen as a defense of a highly questionable inference. If that argument is valid, we are licensed to perform the following inference:

(M1) Since roundness and redness/sharpness are in the same thing, there must be sense data that have these compound properties.

Again, the argument doesn't presuppose this inference. It defends it. But if the argument is good, so is the inference. Again, it is an instance of a more general inferential principle:

(M2) If a thing appears to have compound properties there must be a sense datum that jointly instantiates those properties.

But this leads again to the following embarrassing argument- embarrassing because if the Many-Properties Argument is sound, so is this one. This one can't be; so neither can the Many Properties Argument be:

1. The tomato appears to be red and round *and a tomato*. The dagger appears to be red and sharp *and a dagger*.
2. If a thing appears to have compound properties there must be a sense datum that jointly instantiates those properties. (M2)

So,

3. There must be sense data that are tomatoes and daggers.

We hence arrive at all of the same absurd consequences from this argument that we do from the first. Jackson's own

issues, Jackson, 1977, pp. 131-4).

Finally, let us examine the Colourlessness of the World Argument (CWA). The structure of this argument is somewhat different, and so is the difficulty we will develop for it. The CWA relies on the following inferential principle, embodying what we have called its "extreme scientific realism":

(SR) If a property is not mentioned in physics, it is not a property of physical objects.

Unfortunately for CWA this principle licenses the following argument:

1. If a property is not mentioned in physics, it is not a property of physical objects. (SR)
2. *Being a tomato* and *Being a dagger* are not mentioned in physics.
- So,
3. *Being a tomato* and *Being a dagger* are not properties of any physical objects (follows from (1) and (2))
- So,
4. There are no tomatoes or daggers.

This really is an unfortunate conclusion, both for common sense and for Jackson, and suggests that things have gone deeply wrong. For one thing, we know- if we know anything at all- that there are tomatoes, daggers, and a host of other common physical objects. We might say that we know this with much greater certainty than we know that there are also non-physical, non-mental sense data! Moreover, Jackson's entire motivation for introducing sense data in the first place is to explain our knowledge of such things as tomatoes and daggers. If the consequence of their introduction is that there are not such things, the sense data themselves become rather

argument: (SR) is simply unwarranted. Sciences might be good arbiters of scientific properties. But just as (to mix Shakespearean tragedies) "there are more things on heaven and earth than are dreamed of in your philosophy, Horatio," there are more kinds of things on Heaven and Earth than are dreamed of in your physics, Frank! Commonsense things are among those. (Jackson, 1977, p. 134-9).

But it is now time to ask what the deeper error in Jackson's theory of sense data of which these three fallacies are merely symptomatic. It is this: Jackson confuses the properties of the means by which we represent things with the properties of the things themselves. First and foremost, sense data for Jackson are the mediators of our knowledge of the external world. They explain how it is that we have knowledge of tomatoes, daggers and cats. As we have seen, there are difficulties in explaining how- as they are understood by Jackson- they can do this. But that is what they are intended to do. Now the things we know about do have properties like redness, roundness, sharpness, etc. Jackson is concerned to explain how we know those properties. He attempts to explain that knowledge by positing intermediate phenomena that also have those properties. We know those intermediate phenomena- the sense data- directly, and in virtue of that direct knowledge of them, we come to know external phenomena indirectly. And Jackson is driven then to ascribe these intermediate phenomena precisely the properties that the objects by means of which we know them have. So according to Jackson we know red tomatoes via red sense data; sharp daggers via sharp sense data.

But this is not how mediators of knowledge work. Consider a perfectly ordinary case of mediated knowledge. Suppose I know that there is a red tomato on the kitchen counter right now, not by perceiving it, but through a representation of that fact. Now that representation *might* be red. It might be a colour photograph of that very tomato. (It might even be a tomato- another very similar tomato placed

might be grey- part of a black and white photo of the tomato; it might be transparent- a line drawing of the situation. It might even be acoustic- my wife telling me about the tomato.

The point is that *no properties whatever* of the object of knowledge need be possessed by the states or processes that mediate our knowledge of that object. And even if those mediating states and processes did by chance have the properties of the things our knowledge of which they mediate, that chance sharing of properties would explain *nothing* about how those phenomena mediate that knowledge. Note for instance that the fact that this text is printed in black ink on white paper explains nothing about how it might contain sentences about black and white photographs. (Explaining how representation is possible and what it is for a symbol to mediate our knowledge is itself a hugely complex and controversial affair, and not the topic of this essay.) Jackson's fundamental error then is his attempt not only to locate the properties of external physical objects in knowledge-mediating sense data, but also in attempting to explain that mediating role by reference to the sense data's possession of those properties. No wonder then, that he only succeeds in replicating all of the problems he seeks to explain, constructing sense data that literally must be external objects if they are anything at all, and so knowledge of which must be at least as mysterious as the knowledge they were meant to explicate (Jackson, 1977, p. 130-42).

7. *The Common neo-Cartesian Error and the Demon Cast Out*

I said at the outset that the knowledge argument and the arguments for sense data theory have a common root in a myth that lies deep in our naive self-concept, and that that myth is in turn grounded in a Cartesian picture of human nature. Descartes famously considers the possibility that a

sensations and beliefs that we would have were there actually an external world of the kind we apparently perceive, despite the absence of such a world. He uses this heuristic device to examine the nature of our knowledge, arguing that we might be deceived by such a demon about the external world, in virtue of the fact that our knowledge of that external world is mediated by our knowledge of our own inner mental states, and in virtue of the fact that the demon can disrupt the relation between those states and the external world. On the other hand, he argues, we *cannot* be in error about our own inner states, since we know those immediately. There is no room for the demon to get between us and our own thoughts.

Central to this demonic view of perception and knowledge is the distinction between the *external* world of physical objects and our knowledge of them and the *internal* world of our own inner states and our knowledge of *them*. On the Cartesian picture, implicitly endorsed by Jackson and motivating both of these doctrines, the immediate objects of our inner experience—including our own qualitative and sensory states—are known immediately; are private, in the sense that they can be known only through introspection; and they are non-physical, unlike the objects we know in the external world through their epistemic mediation. (See (Garfield, 1989) for more on this model of introspection.)

This view is very tempting. We can identify at least four reasons for its allure: First, these special inner states and processes seem to provide a necessary bridge between us and the world. Without them, the gulf between ourselves and the objects of our knowledge— between subject and object—appears to be infinite. How else could we cross that chasm without some mediating representations of the kind Jackson posits? Without them we would be locked inside our own minds without any access to the world around us.

Second, this view seems to guarantee and to explain our special relation to ourselves. It explains why our knowledge of our own mind and mental states is so immediate, so immune

special kind of access to our own inner lives that nobody else has to our lives, and that we have to no other domain of knowledge or objects. If the structure of our knowledge and subjectivity is as the Cartesian image would have it, all of this appears to make perfect sense.

Third, this theory of the relation of self to the world appears to explain the possibility of error and illusion as well as the possibility of genuine knowledge. If we have only mediated knowledge of the external world then error simply is what happens when the normal relation between our mediating processes and the external world is disrupted, whether by natural catastrophe or demonic intervention. We can't be wrong about our self-knowledge simply because of its immediacy.

Finally, and perhaps most importantly, the Cartesian-Jacksonian view gives us a very special status in the world. We, unlike insensate physical objects, and perhaps unlike non-rational animals, have non-physical minds, inner arenas in which we are directly aware of non-physical goings on—goings on of the kind that Mary and only Mary could observe when she saw red for that first time, and which stand between us and merely physical tomatoes. This special status is in the first instance perhaps epistemological, but it has distinct metaphysical and perhaps moral and even religious overtones. Without it, and without the epistemology that guarantees it, it might appear that nothing distinguishes persons from the rest of the natural world.²¹

But we should be wary of purchasing uniqueness at such a high metaphysical and epistemological price. The difficulties that beset Jackson's attempts to make sense of this status are symptomatic of at least four deep problems that beset any such account, corresponding to the four motivations just adumbrated. First, in our attempt to construct a bridge between ourselves and the objects of our knowledge, we find that we have launched a regress of bridges. This is because the bridge we create is itself both another object of knowledge,

object. Hence we have two new epistemological chasms where once we had one. We now need a bridge to mediate our knowledge of our ideas and another to mediate the relation between our ideas and the external world. We saw the symptoms of this difficulty emerge as sense data came to look more and more like the objects access to which they were meant to explain, and when we noticed that the possibility of zombies issues in Mary's need to gain some kind of additional and impossible access to her own inner states.

Second—and this is reflected more directly in the zombie problem—the Cartesian picture, instead of explicating our knowledge of others and of the external world through a more transparent self-knowledge, makes that antecedently unproblematic self-knowledge appear as mysterious as it initially conceives knowledge of others and of the external world to be. For as we have seen, if we take this Cartesian picture seriously, our own uniqueness is in part a uniqueness *as opposed to* the mundane character of the zombies. And once the zombies enter our field of possibilities, even as alien beings, our self-knowledge is called into question. Indeed, it turns out that we cannot even know that we ourselves are genuinely sentient, as opposed to voodoo fakes. So what looked like a picture of subjectivity according to which we have a special immediate knowledge of ourselves becomes one on which we do not *even* know our own minds.

Not only does apperception become mysterious, but perception of the external world is rendered doubly mysterious. For not only is it mediated by a veil of ideas that never allows us direct access to the putative objects of our knowledge, but we discover that those representative ideas share none of the qualities of those distal objects. The ideas have all the properties that commonsense attributes to the world—colour, shape, etc. But objects in the world have none of these properties. Nonetheless, somehow we gain knowledge of those objects via these ideas; somehow these ideas represent those objects; and somehow those objects are that to which we

would blush!

So the Cartesian picture to which Jackson implicitly subscribes, and whose consequences he articulates has disastrous consequences for epistemology: Both self-knowledge and knowledge of the external world become apparently impossible, or at least deeply problematic, and it is not at all clear how we could even be epistemically related to either our own mental states or to that which lies beyond. But finally there are metaphysical difficulties in the wings—difficulties that have always plagued Cartesian accounts. Part of the appeal, we saw, of Jackson's updated Cartesian theory of mind is that it makes us appear to be special precisely in virtue of being more than merely physical beings. We have non-physical mental lives and are in direct epistemic contact with a non-physical world. The problem is, however, that this non-physical world and these non-physical mental facts are also, according to this theory—and as they must be in order for it to be at all plausible or for it to explain our perceptual knowledge, as it purports to do—in causal interaction with the physical world. They must be causally related to the physical objects and properties access to which they mediate and to our physical bodies, particularly our sensory and nervous systems. Jackson, of course, grants this.

But this need for a causal anchor to the physical world on each side of the non-physical world raises two problems, one merely methodological and one deeply metaphysical. First, and perhaps least seriously, we should wonder about why an organism would evolve in such a way as to involve an intermediate non-physical link in an otherwise physical process of perception. What process could drive or confer advantage on such an ontological gerrymander? But never mind, perhaps it is some odd spandrel case. More seriously, the physical world appears to be causally closed (at least bracketing quantum phenomena, which are beside the point here). Physical phenomena can in general be entirely explained by reference to antecedent physical causes, and their

effects appear to be entirely physical. There is no evidence or reason to suspect that objects of perception or neural tissues differ from the rest of the world in this respect. Nor is there any account of what such trans-physical causation would look like or what laws might govern it. Absent powerful reasons to suspect such special causal powers and an account of causation that would make sense of them, it appears that Jackson, rather than having provided us with a respect in which human beings are epistemologically and metaphysically singular in any comforting sense, has provided us with an account of our nature that renders us incoherent melanges of physical and spook stuff, whose interaction is at best deeply mysterious. That is to say that he has revived one of the most notoriously inept approaches to the mind/body problem as a side effect of his theory of perception. It is at least ironic that each of the four principal motivations for a view such as Jackson's is in fact rendered all the more inaccessible precisely by a theory of the kind he proposes.

8. *Morals of the Story*

The related difficulties of the knowledge argument and of Jackson's sense data theory of perception, we have seen, both trace to an overwhelmingly tempting vision of the mind as a kind of inner theatre in which plays of ideas are enacted for our private viewing pleasure—plays which, when veridical, grant us knowledge of the external world, and which, when fictional, issue in deception. But regarding the contents of the plays themselves there can be no error, for we are right there in the theatre with them. And furthermore, the stuff of these dreams is utterly discontinuous with the stuff of nature. Otherwise, it couldn't make it into this non-physical theatrical space of our minds. When put this baldly, no-one, not even Jackson, would sign on for such a theory of human nature. (And indeed, such a theory of human nature has been roundly castigated over the past forty years any number of times. See

the most recent and popular version of this casting.)

Despite such regular exposure and refutation, however, this image has a perennial attraction. The ease with which a philosopher as subtle and sophisticated as Jackson is seduced by this vision is testimony to its allure and to the depth of its sedimentation in our collective psyche. We are indeed possessed by the Cartesian demon, and its operations in philosophy and in everyday life are maddeningly occult. Only by rigorous examination of the views we find most plausible do we discover its insinuations into our otherwise up-to-date, even postmodern explicit conceptions of ourselves. And as we have seen, the image the demon urges upon us is pernicious indeed. Despite the Mephistophelean promise of human uniqueness and of knowledge assured, the Faustian bargain into which we are thus implicated is one that costs us even knowledge of our selves, and any hope of a coherent story of what it is to be a human being in a physical world.

But there is a final, more positive moral to the story: The philosophical clarity and rigour of a philosopher such as Jackson allows us to see clearly, through focusing so precisely on the details of phenomena like perceptual knowledge and the qualitative aspects of perceptual states just where these deeply sedimented myths intrude into our explicit self-concept. This is the very engine of philosophical progress. Without such analysis pursued with Jackson's relentless and intelligent rigour, it is far too easy to succumb to the misplaced conceit that all of the demons of our Cartesian past have been exorcised, and in that false confidence to lapse unconsciously and unwillingly into those very Cartesian misconceptions and conceits. In fact, we discover through examining analyses such as Jackson's that they are actively at work in our most plausible and natural intuitions regarding our own nature. We then have the opportunity for a more thorough ritual of exorcism (though to call it final would surely be an act of philosophical hubris).

1 philosophical progress is understood as a structure. Only the clearest and most forceful articulations of a view allow us to see its deep presuppositions. Only examination of those presuppositions allows philosophical dialectic the depth we have a right to expect, issuing not only in a clearer understanding of some detail of the philosophical terrain, but in an entirely new perspective on the landscape. We hence owe a great deal to Jackson for making so explicit so much of what was heretofore implicit in our pre-reflective self-image. My critical remarks in this essay are hence offered with the greatest respect for Jackson's genuine philosophical contribution.

1 This essay is dedicated to the memory of Prof Thomas Tymoczko, who insisted on and exemplified clarity of thought, complete intellectual honesty and rigour, and a commitment to enthusiastic, joyful and constructive philosophical debate, the goal of which was always the truth, and which always reflected a genuine openness to the views of his interlocutors and a concern that all involved profit from the interchange.

2 I thank Prof Max Cresswell for helpful conversations on Jackson on sense-data. Much of my discussion of that subject is influenced both by his fine essay on that material and by our conversations. I also thank Mr J.C. Beal for a spirited e-mail interchange on the knowledge argument, for access to some of his fine draft material on this topic and for very useful comments on initial formulations of the zombie argument. My views and formulations of several arguments owe a great deal to his insights and responses. I also thank the Propositional Attitudes Task Force, especially Lee Bowie, Murray Kiteley and Ernie Alleva for detailed comments and criticism and Prof Frank Jackson for useful comments on an earlier draft and for an exchange which clarified both his current views (which differ somewhat from those defended in the texts I discuss) and the differences that remain between us. Thanks especially to Prof Frank White whose comments on the materialism of David Armstrong catalysed my sense of the core of my own

disagreement with Jackson and whose vigorous opposition to my own views on the philosophy of mind has forced me to much greater clarity. Prof White will probably agree with almost nothing I say in this essay. But almost every line is in some way a response to his powerful critique. And thanks to Dr Moira Nicholls for helpful editorial suggestions.

I should note my complete agreement with this view. See (Garfield, 1988).

But see also (Conee, 1994) for similar replies as well as some new ones.

Intentional contexts include those that fall under the scope of psychological verbs like "believes," "knows," "desires," etc. The verbs are intentional because they denote psychological states that *intend* or contain contents—in the case of these verbs, propositional contents like "that Queen Elizabeth II just walked by". These verbs create contexts that are notoriously also *intensional*. That is, in such contexts, the substitutivity of co-referring terms is impossible, *salva veritate*. So, while if Paul touches Queen Elizabeth II, he *ipso facto* touches the richest woman in the world. But though he may desire to meet the Queen, he may have no desire to meet the richest woman in the world (believing her to be someone else).

The idea of using zombies so understood against the knowledge argument is inspired by Dennett's use of zombies in (1991) and by his use of a zombie argument against Michael Lockwood in a symposium on consciousness in which we all participated at Amherst College in 1992. In fact, my zombie argument can be read as a specific and explicit application of some general suggestions he makes in that book (see especially pp. 404-405) to the knowledge argument. On the other hand, whereas Dennett argues not only that zombies are possible but that *we are* zombies, I argue that they are conceptually impossible. It is curious that these views turn out to be close cousins. The crucial point is that zombies and real humans are not *compossible*.

Lee Bowie argues (e-mail correspondence) (1) that beliefs are qualitative states; (2) that even if they are not, it is a necessary condition of having beliefs that one be the subject of qualitative states; and (3) that even if it is not, having *qualitative beliefs* requires being the subject of qualitative states. He argues that

by assuming that beliefs are non-qualitative I beg the question against the Knowledge Argument.

With regard to (1): First point: I find the claim utterly implausible on its face. After all, think of how many beliefs you have (no doubt infinitely many). Is there something that it feels like to have each of them? Could it really be the case that different people with the same belief thereby feel some particular way? When you change your mind about a belief, do you change the way you feel? What does it feel like to believe that the number 19 is odd? Second point: The zombie argument in any case is a reductio argument against the Knowledge Argument. Jackson takes it for granted (as do I and virtually everybody else) that beliefs are not qualitative states. So if they turn out to be qualitative states the Knowledge Argument already fails. I at least grant Jackson this premise and show that the argument *still* fails. Third point: If Jackson and I are both wrong about this, and Bowie is right then zombies are impossible. But I am about to show that if the Knowledge argument succeeds then zombies *must* be possible, *and* that they are *in fact impossible*. So if Bowie is right, my reductio in fact succeeds.

So if Bowie is right, my reductio in fact succeeds.

(2) is another way of saying what I am about to conclude, *viz.*, that zombies are impossible. It hence represents no criticism of my argument. For a reply to (3) see note 8, below.

One might respond at this point that zombies have no qualitative beliefs. After all, they have no qualia. So, how could they have beliefs about qualia? (This objection was suggested to me by J.C. Beal in an e-mail interchange.) Note, however, that to respond in this way begs the question pretty seriously. It cleaves off a set of beliefs that we can have but that zombies can't. But zombies are supposed to be like us cognitively in every respect. Why deny them *these* beliefs? Because of the entailment regarding the impossibility of self-knowledge? That would be to argue in a rather tight circle. Moreover, given the seamlessness of the web of belief, and the multiple inferential links between qualitative and non-qualitative beliefs, to deny zombies *these* beliefs would be to deny them a host of other related beliefs. It would be impossible to circumscribe the set of beliefs inferentially connected to qualitative beliefs. Hence to deny that zombies could have qualitative beliefs would be to deny that they could share any of our beliefs, which would be in turn to deny the

possibility of zombies so described, which would be to concede the conclusion of the zombie argument.

In particular, it vindicates objections three, five, six and seven. I thank J.C. Beal for this disjunctive formulation of the conclusion.

The Propositional Attitudes Task Force has raised a legitimate worry here concerning my understanding of what a fact is. There are notorious philosophical problems associated with specifying what constitutes a fact, or when two putatively different facts are in fact the same fact. I am quite self-consciously sidestepping these controversies. In this context that is legitimate, as I am demonstrating the incoherence of Jackson's position. It is he, not I, who claims that there are these additional facts. I merely claim that *whatever his account of facts* he gets no argument for the existence of two domains of such things—a physical one and a mental one (again, no matter how the latter two domains are characterised). The burden or proof—and the burden of theoretical exposition—is hence in this debate squarely on Jackson's shoulders. I even (generously) grant him whatever account he may like of facts, and of the physical and mental domains, and show that the Knowledge Argument cannot be an argument for special non-physical facts.

The isolation of the problematic status of such putatively private objects of introspective consciousness is due originally to Wittgenstein in (1956, §§ 208-318). The difficulties are amplified and sharpened by Sellars in (1963a, §§1-7, 32-38)

As such it is a descendent of the argument of Wittgenstein (1956, §§ 208-318) and of Sellars (1963a, §§ 32-38, 47-48)

For a discussion of another, more empiricist version of the myth and its refutation, see (Garfield, 1989).

Contrast this with the conclusions drawn by Churchland (1985) But this analogy should not be pressed too far: John may properly be said in one sense to have two objects of knowledge: QE II's regal status and her bank balance. That is not the analogy we are after here. Rather we are comparing his two different modes of knowledge of one woman to Mary's two idea of "two different modes" is not obvious, and there must

be some way of doing this. Jackson and the members of the Propositional Attitudes Task Force have urged this point forcefully in correspondence. In particular, if I am right, it cannot require that we simply specify different descriptions under which a thing is known, or different properties it is represented as having. I don't have a view about the right way to spell this out. But it does nonetheless seem clear that whatever one's ontology of facts, it must be possible to have different cognitive relations to them, and I suggest that first-person/third-person captures in a plausible way that difference in this case.

For expositions of the classical theory see (Ayer, 1940; Broad, 1925; Moore, 1953; Pitson, 1985; Price, 1932). For the critique that buried it until its Jacksonian revival see (Sellars, 1963a; Sellars, 1963b).

It may be misleading to say that it "thus appears oval." It certainly doesn't *appear to be* oval. But it is no better to say that the appearance is oval, as some would prefer. That seems to rely appearances and hence to prejudice our analysis. The most accurate way to put it—though hardly felicitous—is to say that its image on the retina is that that would be projected by an oval object viewed face-on. I thank Lee Bowie for pointing this out.

I am deeply indebted in my exposition of Jackson's central arguments for the sense data theory and in my first level critique of these arguments to (Cresswell, 1980). Indeed, I adopt Cresswell's exposition and critique virtually wholesale, with some slight modifications in the exposition for the sake of clarity. I then turn that critique in a slightly different direction when it comes to my diagnosis of the deeper difficulties with Jackson's view. I also note that the structure of Cresswell's critique of Jackson's particular formulation is anticipated in Kiteley's (1960) diagnosis of the fundamental error in sense data theories.

For a discussion of these issues, see (Churchland, 1979; 1985) and (Garfield, 1988; 1989; unpublished).

I thank Prof Frank White for emphasising the centrality of this issue to neo-Cartesian views of the mind.