1 Making Consciousness Safe for Neuroscience

Andrew Brook

Work on consciousness by neurophilosophers often leaves a certain group of other philosophers frustrated. The latter group of philosophers, which includes people such as Thomas Nagel, Frank Jackson, Colin McGinn, Ned Block, and David Chalmers, believe that consciousness is something quite different from the brain circuitry or other processes that are active in cognition. They feel frustrated because work on consciousness by neurophilosophers usually ignore this view, yet proceed from an assumption that it is wrong. This work tends to assume, simply assume, that neuroscience will not only identify neural *correlates* of consciousness (which virtually all parties to the current consciousness debate now accept), but (perhaps with the assistance of cognitive science) will eventually tell us *what consciousness is.* That is to say, it assumes that consciousness simply is a neural/cognitive process of some kind. Even more, it assumes that consciousness is a neural/cognitive process similar in kind to the processes that underlie (other aspects of) cognition and representation. That is to say, it assumes that consciousness is an aspect of general cognition.

Consciousness has appeared to be weird and wonderful to many people for a very long time. Daniel Dennett captured the feeling very nicely many years ago.

> Consciousness appears to be the last bastion of occult properties, epiphenomena, immeasurable subjective states – in short, the one area of mind best left to the philosophers. Let them make fools of themselves trying to corral the quicksilver of "phenomenology" into a respectable theory. [1978a, p.149]

Consciousness no longer appears *this* strange to very many researchers but the group of people just mentioned continue to hold that it is very different from any brain or other process active in cognition. By contrast, like most consciousness researchers now, neurophilosophers simply take for granted that consciousness will be domesticated along with the rest of cognition – indeed, that it will turn out to be simply an aspect of general cognition.

I am sympathetic to the position assumed by these papers. However, I do not think that one can simply ignore the opposition. In my view, one must confront their arguments and show where they fail. In addition to the general desirability of not ignoring one's opponents, there is a specific reason why this needs to be done in the case of consciousness. If the opposition is left unchallenged, it can easily appear as though the neurophilosophy of 'consciousness' is in fact not talking about *consciousness,* has subtly changed the topic. It can easily appear that these papers are merely talking about correlates of consciousness, not real McCoy, consciousness itself. The opposition does not have much to say about what this something else might be like – a deep streak of what Owen Flanagan (1991) calls mysterianism runs through this work – but they think that they have strong arguments in favour of the claim that it *is* something else. I think these arguments must be answered, not ignored – which is what I will try to do in this paper.

The arguments all have the same cast. They consist of thought-experiments designed to show either that complete cognitive functioning, even cognitive functioning of the highest sort, could proceed as it is in us without consciousness, or that it could proceed as it is in us even if the contents of consciousness were very different from ours. The most common arguments of the former kind are zombie thought-experiments: something could behave like us or function cognitively like us or even be a molecule-for-molecule duplicate of us without being conscious. One common form of argument of the second kind is exemplified by inverted spectrum thought-experiments: where something seems green to us, it could seem red to another. However, because this other has been trained to express its

experience using the word 'green', and so on, its behaviour – and even its cognitive functioning – could be just like ours. These are the kind of arguments that I propose to go after.[1]

Broadly, there are at least two ways in which one might go after such arguments. One way would be to show that the opposing point of view is unproductive. We will do that. Indeed, I will suggest that it is so unlikely to shed light on most of the interesting aspects of consciousness that consciousness simply could not have the character that the view says it has. The other would be to show that the arguments do not succeed. I will suggest, indeed, that there is deep incoherence built into most of them.

Thus the paper has three parts. First we will articulate the difference between the two points of view in detail. Along the way we will examine one respect in which most neurophilosophy fo consciousness is different from even physicalist philosophy of consciousness, much of it at any rate, and we will examine and reject the old charge that most neurophilosophers are really eliminativists about consciousness. Then we will do a (by the nature of the current science necessarily provisional) assessment of the prospects of the anti-physicalist point of view as an explanatory theory of consciousness. Finally, we will sketch a general line of argument that, I will argue, undermines most of the thought-experiments invoked in support of anti-physicalism.

As will have become clear, this paper won't have a lot of neuroscience in it (though some of it appears in the final section). One of philosophy's important roles historically has been clearing conceptual ground, removing confusions and confounding notions, providing empirical researchers with concepts adequate to do their work with clarity and precision. There is still a lot of obscuring underbrush in consciousness research. The profusion of current terminology alone is enough to show this. There is access consciousness, phenomenal consciousness, self-consciousness, simple consciousness, background consciousness, reflective consciousness, creature consciousness, state consciousness, monitoring consciousness, consciousness taken to be coextensive with awareness, consciousness distinguished from awareness, qualia, transparency, consciousness as higher order thought, higher order experience, displaced perception ... and on and on. The aim of this paper is to clear a small part of the underbrush out of the way.

1. The two approaches to consciousness

Whatever their position on other issues, most researchers now take at least one central form of being conscious of something to consist in it being like something to represent that state, and conscious states to be states that make us conscious in this way. When I consciously perceive the words on my

---

1. Chalmers' well-known (1995) distinction between what he calls the easy problem and the hard problem of consciousness starts from this distinction between the cognitive role of representations and something appearing to be like something in them. Understanding the former is, he says, an easy problem, at least compared to understanding the latter. The easy problem is to understand the inferential and other roles of such states. The hard problem is to understand how, in these states or any states, something could appear as something to me, how certain stimulations of the retina, processing of signals by the visual cortex, application of categories and other referential and discriminatory apparatus elsewhere in the brain can result in an *appearing,* a state in which something *appears* a certain way. Chalmers says that the easy problem is easy because it is simply the problem of the nature and function of representation in general, while the hard problem is hard because it is *sui generis*, quite unlike any other problem about cognition that we face. If the first problem is easy, I'd hate to see what a hard one is like but on anti-cognitivism, the two will at least be quite *different* problems.

computer screen, it is like something to perceive those words, in Nagel's (1974) now-famous phrase, and the perception is a conscious state. If I shift attention to the perception itself ('How clearly I can see print on this screen?), it then (at least then) becomes like something to have, too – which makes the case for calling it a conscious state even stronger.[2] Without committing ourselves as to whether these are the only forms of consciousness, let us take these phenomena and their common core, it being like something to have them, as our target. As we have seen, researchers divide over whether consciousness so understood (from now on simply 'consciousness') is a kind or aspect of cognition or of the brain processes active in cognition.

By contrast, as we have also seen, others say, or assume, that consciousness *does* consist of a kind or aspect of cognition and the brain processes active in cognition. This, in fact, is the position of virtually all empirical investigators, whether cognitive or neuro- scientists (Baars, Posner, Mack and Rock, Shallice, Jackendoff, Newman, Zeki, Goodale and Milner, Crick and Koch and many others), and some philosophers (Dennett, Rosenthal, Tye, Dretske, **Prinz**, Thompson, P. M. and P. S. Churchland, and many others).

The two sides of this division are sometimes called representationalism and anti-representationalism. The name is not entirely apt. Those who are hostile to the idea that consciousness is representational would be equally hostile to an idea that it is some other cognitive property. And on the other side, there are those who view consciousness as a perfectly straightforward aspect of cognition but reject the idea that consciousness is a *representational* phenomenon. Representations (or, if you reject the idea of representations, your favourite surrogate for representations[3]) are only one part of a cognitive system. There is also all the machinery for processing representations (and the emotions, and volitions, and other things). Theorists who hold that consciousness is an aspect of cognition but not a representational aspect include, for example, all the researchers who hold that consciousness is attention or a kind of attention. So let us instead call the two sides of the division *cognitivism* and *anti-cognitivism* about consciousness.

> **Cognitivism** – the view that consciousness is a representational property of representations or a cognitive property of the system that processes representations,

and,

> **Anti-cognitivism** – the view that consciousness is neither a representational property of representations nor a cognitive property of the system that processes representations,

2. Another divide in research on consciousness

Another divide in recent work on consciousness has not been as prominent as the cognitivist/anticognitivist one. It will help us situate neurophilosophy vis-à-vis the opposition. Many

---

2. Confusion and (often implicit) disagreement abound in the use of consciousness terminology and there is some stipulation in what I have just said. I hope the result is nonetheless not controversial.

3. So long as you allow that we need some term or set of terms for talking about how things appear to people, I don't care whether you call it representation or something else. Note that nothing here hangs on a particular view of representation. Whether one views being a representation as a matter of referring to, standing in for, reliably covarying with, being semantically evaluable, or something else, anti-cognitivists will deny that consciousness has that character. And likewise for those who view consciousness not as a property of representations but of whole cognitive systems.

philosophers of consciousness focus on individual psychological states – individual perceptions or feelings or imaginings (Chalmers 1996; Tye 1995) – or at most tiny combinations of such states (a thought directed at an experience, for example; Rosenthal 1991). Let us call this the *atomistic approach* to consciousness.

> **Atomist approach to studying consciousness** – the view that conscious states can be studied one by one or in small groups without reference to the cognitive system that has them.

Atomists about consciousness talk about conscious states one by one ('what is it like for something to look red?'), or at most in tiny groups, for example a thought directed at a perception (the so-called higher-order thought view [Rosenthal 1991]), and ask questions such as, 'When a state is like something to have, what is this aspect of the state like?' The answers to this question then split along the lines of the cognitivist/anticognitivist division. Almost the whole of the massive literature on *qualia* is atomistic in this way. ('*Qualia',* a philosopher's term, is a term for the felt quality of an individual conscious state, 'what it is like to have it'.) But notice: atomists of either stripe ignore the cognitive system whose states these states of consciousness are. They may say the words, '... look red *to me'* but they do nothing with the addition.

Most experimentalists by contrast focus on properties of whole cognitive systems: global workspace (Baars 1988), intermediate level of processing (Jackendoff 1987), attention (Posner 1994; Mack and Rock 1998), phaselocked spiking frequency (Crick and Koch, 1994), or something similar. Let us call this the *system approach* to consciousness.

> **System approach to consciousness** – the approach to consciousness that views it as a property of whole cognitive systems, not individual or small groups of representations or properties of individual representations such as their *qualia* (their 'felt quality').

For Posner or Mack and Rock, for example, to be conscious of something simply is to pay attention to it. Here is Mack and Rock: "Attention [is] the process that brings a stimulus to consciousness" (Mack and Rock 1998), "if a ... percept captures attention, it then becomes an explicit percept, that is, a conscious percept" (Mack 2001, p. 2). Posner (1994) captures the spirit of this line of thinking about consciousness nicely:

> an understanding of consciousness must rest on an appreciation of the brain networks that subserve attention, in much the same way as a scientific analysis of life without consideration of DNA would seem vacuous. [Posner 1994, p. 7398]

Now, what is interesting about this atomist/systems divide is that, unlike a great many other philosophers, virtually all neurophilosophers are system theorist. Dennett's (1991) multiple drafts model is a prime example. For him, consciousness is a matter of one or more of the multiple drafts of various descriptions and narratives in us achieving a certain kind of dominance in the dynamics of the Pandemonium-architecture of cognition. (Curiously, he says almost nothing about attention.) Paul Churchland is another example. Here is how Churchland summarized his approach recently:

> [Consider] the brain's capacity to focus attention on some aspect or subset of its teeming polymodal sensory inputs, to try out different conceptual interpretations of that selected subset, to hold the results of that selective/interpretive activity in short-term memory for long enough to update a coherent representational 'narrative' of the world-unfolding-in-time, a narrative thus fit for possible selection and imprinting in long-term memory. Any [such] representation is ... a presumptive instance of the class of *conscious* representations. [Churchland 2002, p. 74]

In this volume, **Prinz** clearly takes a system approach. Indeed, the system approach to consciousness once dominated in philosophy – think of Descartes and Kant.

Of course, the system theorists we have just considered are all system cognitivists. Anti-cognitivists, however, can also be system theorists. Indeed, of the five anti-cognitivists that we

mentioned at the beginning, two of them are broadly atomistic (Chalmers and Block) but two of them take a broadly system approach (Nagel and McGinn).[4] For the atomist anti-cognitivist, individual conscious states are lined up one-to-one with individual representations. However, what makes a state a conscious state is nothing representational, not even when a representation is a conscious state. Representations could do the representational work they do in the absence of consciousness or in the presence of very different conscious content.

System anti-cognitivists come by their hostility to cognitivism in a variety of ways.[5] Many system 'anti-cognitivists' are philosophers. Nagel (1974), for example, argues that the only way to understand what a point of view is, is to have one. The more one tries to adopt a third-person or impersonal point of view on a point of view, the more one moves away from what it is. There is no reason, however, to think that the same restriction to the first-person governs our understanding of representation or other cognitive functioning. If so, there is reason to doubt that consciousness is representational. McGinn (1991) argues that we cannot know how consciousness is linked to the representational activities of the brain; consciousness is 'cognitively closed' to our kind of mind the way that physics is to field mice. If we expect a scientific understanding of cognition to be possible, this would be a reason to group him with the anti-cognitivists.[6] Among non-philosophers, many different anti- or at least non-cognitivist views can be found. Perhaps Penrose's (1999) view that consciousness is a quantum phenomenon rather than anything active in cognition is perhaps typical.

With this, the issue is now clear. When neurophilosophers simply assume that consciousness is cognitive, some aspect of the machinery for managing representations, it will strike anti-cognitivists of both stripes, both atomist and system anti-cognitivists, that whatever they are talking about, it isn't consciousness. Cognition and representation of all kinds could proceed without consciousness, and there is no reason to think that consciousness is anything like either of them.

We should note before we leave the system approach that there is enormous diversity in views that could be regarded as taking this approach. System approaches are far from a monolith. We cannot begin to explore the whole range of options here but here are some of the leading views. Consciousness consists in a global workspace of a certain kind (plus some other things) (Baars 1988), Consciousness is an intermediate level of representation, a phonetic or similar level between acoustic or visual input and full-blown conceptual content (Jackendoff 1987). In understanding consciousness, attention should be singled out for special ... attention. (Posner 1994; Mack and Rock 1998). Consciousness is attention feeding working memory (**Prinz,** this volume). Consciousness is the result of multiple constraint satisfaction, a property or properties emergent on brain/world interactions. Consciousness is a form of self-organization in a dynamic system. Consciousness is a draft winning the competition for cognitive resources (Dennett 1991). Consciousness is, or is a product of, phaselock synchrony (Crick and Koch 1994). Consciousness is the result of a certain tensor phase-space processing (Churchland 1995). All this from people who broadly agree with one another! There is no agreement even on something as basic as whether consciousness is a biological property (P. M. Churchland 2002) or a cultural/information-engendered property (Dennett 1991). The system approach

---

4. Jackson is difficult to classify.

5. I am not sure that their authors would all accept the label 'anti-cognitivist' but I think that it is a fair label.

6. This consideration applies to Nagel, too. Since science is done from an impersonal point of view, Nagel is also arguing for a limitation on how far we are able to understand consciousness scientifically. What a point of view is like will elude science.

to consciousness is in a considerable mess.

3. Eliminativism: A Third Axis?

Of course, one reason for leaving consciousness out would be a belief that there is no such thing and one form of the charge against some neurophilosophers is that they urging that there is no such thing as consciousness when there patently is. Thus, it may seem that we need a third distinction, between eliminativists and non-eliminativists about consciousness. Here is what I mean by 'eliminativism':

> **Eliminativism about consciousness** – the view that the term 'consciousness' will prove not to be a theoretically useful term, that nothing exists that resembles what we take consciousness to be like.

Well, if the views of any neurophilosophers entail that consciousness should be eliminated in favour of something else – something infinitely less interesting and important, of course – such a result would be entirely inadvertent. There are very few deliberate eliminativists about consciousness.

At one time, Patricia Churchland (1983) and maybe Paul Churchland flirted with the idea. However, even at their most eliminativist, they never advocated wholesale replacement of our consciousness talk in the way that they did for our intentional talk. In recent years, they have backed away from eliminativism about consciousness almost entirely. As we saw in the previous section, Paul Churchland is now quite happy to talk about consciousness as a perfectly real phenomenon in need of scientific exploration.

Some think that Dennett's (1991) multiple drafts model of consciousness is eliminativist. This would be quite wrong, in my view. Dennett certainly rejects a dominant way of thinking about consciousness, what he calls Cartesian materialism. But to reject a *theory* of consciousness is not to deny the existence of consciousness. To the contrary, Dennett has said repeatedly that consciousness is a perfectly real phenomenon (1998, pp. 135, 146). As he sees it, consciousness involves more interpretation by the cognitive system than has been thought, a system that in turn has less unity and stability, less universal general cognitive structure than has been thought, and the resulting conscious states have less determinability and temporal stability than has been thought. However, none of this is to deny that there is something appropriately called consciousness. Dennett himself says that he wants to be a deflationist about consciousness, not an eliminativism: he wants to deflate the pretensions of theories that insist on seeing consciousness as something weird, wonderful, and exotic (2000, pp. 369-70).

And there are good reason why eliminativism about consciousness is extremely rare. Could 'consciousness' turn out to be a theoretically useless term? That would require that the term is merely a misleading name for a variety of processes much better named and described by other terms, so that 'consciousness' is a vague umbrella term for a diverse group of different things more perspicuously dealt with by giving each its own name (P. S. Churchland 1983). Or that nothing in us is much like what the term 'consciousness' depicts pretheoretically, there is nothing in the brain that could be usefully labelled 'consciousness'. (We don't yet have anything remotely resembling a story about what 'consciousness' is *supposed* to name, so exactly how one would determine this is a nice question.)

Both ideas mislocate the role of the term 'consciousness' and cognates in our cognitive life. Unlike, say, 'intentionality', 'consciousness' is not a term of art. The notion of consciousness has deep roots in everyday discourse. We talk about losing and regaining consciousness. We talk about becoming conscious of this and that. We talk about being intensely conscious, for example of oneself. It is unlikely on the face of it that all these modes of discourse rely on an implicit theory, or are describing nothing real or a bunch of things better discussed in a different vocabulary (thought

doubtless some of the latter will turn out to be the case).

In short, neurophilosophers take consciousness to be a perfectly real phenomenon, just as real as their anti-cognitivist opponents take it to be, and they want their theories to explain *it,* not something picked out by some successor notion.

4. Prospects for Anti-cognitivism

As we said earlier, one way to undermine anti-cognitivism would be to show that it is so unlikely to shed light on interesting features of consciousness that consciousness could not have the character that the view says it has. So let us ask, does anti-cognitivism have any serious prospect of explaining interesting features of consciousness? When compared, for example, to system cognitivism, which has the better prospect?

What do we want a theory of consciousness to explain? As has often be said, consciousness:

can be faint, full, etc.

can be independent of, indeed can continue in the absence of, sensory inputs.

disappears in deep sleep, and . . .

reappears in dreams.[7]

Then there is consciousness of self. On the face of it,

Consciousness of oneself and consciousness of one's acts of representing, desiring, and so on seem to be two different things.

Moreover,

Consciousness of self and the cognitive activities that yield it appear to have some unusual properties. Consciousness of self seems to use what Shoemaker (1968) called reference without self-identification, the resulting consciousness seems to have what he called immunity to error through misidentification with respect to the first person, and the use of first-person pronouns seems to be, to use Perry's (1979) term, essential.

Next, consider the conscious cognitive system. There has to be such a system; consciousness is a matter of *something being conscious* of something.

Consciousness requires a conscious subject. (Atomism falls short right here.)

What is a system capable of consciousness like? Here are some features of such a system:

Such a system has some general cognitive features:

Often how things appear to such a system is the result of cognitive activity, sometimes intense activity, on the part of the system.

Many of the global cognitive faculties of such a system are closely linked to consciousness, memory, for example, attention, and language.

For consciousness, a system simply having information as a result of representing this, that or the other is not enough; the system must make cognitive use of the information.

Consciousness requires a system that is capable of representing; there is a representational base

7. This list of four items and the items in the second list below are derived from Churchland's (1995, pp. 213-14) list of the Magnificent Seven features of consciousness. I go beyond his list in a number of ways.

to consciousness.[8]

Usually a cognitive system is conscious of whole groups of represented items in one 'act of consciousness'. (Brook and Raymont, 2000)

Usually when a cognitive system is conscious of groups of items in one act of consciousness, it is also conscious of representing them and of itself as the common subject of these representations.

Explaining these features of consciousness is a basic requirement on a theory of consciousness. How do the different approaches do?

The answer is straightforward. When it comes to explaining this list of features, both for atomistic and system anti-cognitivist just claw the air.[9] By contrast, cognitive system approaches at least holds out a hope of being able to explain some of these features of consciousness. If so, it is unlikely that anti-cognitivism is talking about consciousness.

Anti-cognitivism is impotent in another way, too. An adequate theory of consciousness should be able to explain the all the main kinds of consciousness. Two of them are,

Consciousness of the world – being conscious of the world around us,

and,

Consciousness of oneself and one's states – the consciousness that we have when, for example, we are conscious of *representing* items in the world or conscious of *ourselves* representing items in the world,

Most (maybe all) variants of anti-cognitivism have nothing to say about consciousness of the world. In zombie thought experiments, the organism's relationship to its world is meant not to change. The question of whether they would still be conscious of the world is seldom asked but, once asked, it is far from clear that they would not be. For Nagel, it is what it is like to have a point of view that is forever beyond the reach of science, not what a point of view is. Though inverted spectrum thought experiments are about how the world appears to me, not how something in me appears to me, they are not really about consciousness of the world either. What is supposed to appear inverted is something purely internal. *Ex hypothesi,* the discriminations and comparisons that I make, the actions I launch, all this and everything else to do with my colour relationships to the world around me remain the same. These thought experiments take it that appearings can change while nothing else changes, so they too have to be focussing on a property of consciousness of representations, not consciousness of the world. And so on.[10] So much for the idea of a unified theory of consciousness.

In this section, we have shown the potential of anti-cognitivism as an explanatory theory of consciousness to be fairly dismal. But there are still the arguments for anti-cognitivism.

---

8. This statement is not the same as saying that consciousness is representational, or even that consciousness of something requires that we be representing it. All it says is that consciousness requires *a system* that can represent. As we noted earlier, some kinds of conscious states may not be representations, mood states for example, or mystical states. (Actually, I think that all such states are representational but cannot argue the point here.)

9. To be sure, for some mysterians, this is a cause for rejoicing, not regret.

10. Some theorists make a distinction between state consciousness, those states of our that are conscious, and creature consciousness, which is at least something like our consciousness of the world. They then focus in their theory of consciousness on the former. This illustrates the deficiency we are describing very nicely.

5. The frustration and arguments of anti-cognitivists

The frustration

We have said that for the anti-cognitivist, cognitivist system approaches to consciousness are not talking about *consciousness* or they miss the most interesting and central features of it. A passage from Dennett illustrates how the frustration can arise. Says Dennett, "We are beginning to discern how the human brain achieves consciousness. [I and others] see convergence coming from quite different quarters on a version of [Baars'] global workspace model" (2001). Statements like this tend to make anti-cognitivists crazy! Why? Because it seems perfectly easy to imagine a global workspace grinding away doing its thing with no consciousness at all. The point can be generalized. For any form of representation and any representing system, couldn't one imagine such a system doing all the wonderful cognitive things that it does without consciousness? If there is anything to this challenge, then consciousness is not representational or anything cognitive. So it cannot simply be ignored.[11] Instead, we need to examine whether there is anything to it.

The arguments

Here is one way in which an argument for this anti-cognitivism can get going. When something appears to us to be a certain way, the representation in which it appears can play two roles in our cognitive economy. The contents of the representation (or even the representation itself) can connect inferentially to other representations: if the stick appears to have two straight parts with a bend in the middle, this will preclude representing it as forming a circle. The representation can also connect to belief: if the stick appears straight with a bend in it, I will not form a belief that it bends in a circle. And to memory: I can compare this stick as it appears to sticks I recall from the past. And to action: if I want something to poke into a hole, I might reach for the stick. In all these cases, so long as I am *representing* the stick in the appropriate way, it would seem to be irrelevant whether I am *conscious* of the stick or not. But I am also *conscious* of the stick – it *appears* to me in a certain way. Now, it seems at first blush plausible to say that my representation could do the representational jobs just delineated whether or not I was conscious of the stick or of my representation of it. If so, we might begin to suspect that being conscious plays no independent representational role.

The anti-cognitivist then advances arguments aimed at turning this suspicion into a conviction. The best known is the zombie thought-experiment introduced earlier. There could be creatures just like us behaviourally, cognitively, or even physically who nevertheless are not conscious.[12] Though they

---

11. Dennett has been fighting in the consciousness wars for too long to neglect the opposition himself. Indeed, in the very paper just cited he says that he will "diagnose some instances of backsliding and suggest therapeutic countermeasures."

12. For a good sample of this literature, see the *Journal of Consciousness Studies* target article by Flanagan and Polger (1995) and the remarkable array of comments that it generated. Inverted spectrum and a host of other thought experiments (including dancing qualia, inverted earth, shrinking brain, and expanding brain) raise similar questions but here we will confine ourselves to zombie thought-experiments. See also Polger's (2000) followup article and Dennett's (2000) reply. Note that zombie

are built and behave in ways wondrously like us, all is 'dark' inside.[13] Zombie thought-experiments seek to establish that a representation is one thing, what makes it a conscious state (its *qualia)* is another. If this thought-experiment establishes that a split between cognition and consciousness is so much as possible, then all forms of cognitivism about consciousness are in trouble.[14]

A second familiar argument for this "neo-dualist" conclusion, as Perry (2001) calls it, is the old thought-experiment about inverted spectra: the way in which colours appear to me could be inverted with respect to how they appear to you without changing how our respective representations of colour function as representations?[15]

A third argument could be advanced, too, though it has seldom been used. We will call it the argument from *imprisoned mind*s. An imprisoned mind is a mind that is working perfectly well but cannot express itself in behaviour (for many people, a scenario of utter horror). Unlike zombies, imprisoned minds actually occur. Curare can produce imprisoned minds, for example.

Curare is a muscle paralysant and in sufficient doses produces total paralysis. Not even an eyebrow moves. It is added to some anaesthetic mixes to keep surgery patients from twitching and moving. For a time, more curare as a ratio of the whole and lower doses of central nervous system (CNS) suppressors were used, especially with children. Upon regaining consciousness, some of these patients appeared traumatized, so a surgeon who had to have a small procedure done volunteered to undergo the surgery with the same anaesthetic mix. To his horror, he did not lose consciousness. He just lost all capacity to express what he was feeling. So he felt every slice of the scalpel, the insertion of every stitch – and he could do absolutely nothing about it![16]

Imprisoned minds also arise as a result of strokes. Such a stroke can completely paralyse the body but leave the patient conscious for a few days. Many of these patients retain control over eye movements but it is suspected (and could be only suspected: this is a real life problem of knowledge of other minds) that some patients lose even this minimum channel of communication and are completely unable to communicate. Fortunately, such patients die after only a couple of days (so far as we know?) (Zeman 2003; see also Calvin's 2003 review).

There are also cases in which victims of accidents who are in a persistent vegetative state have

---

thought-experiments have extremely broad scope; they aim to establish that consciousness could be absent from *anything* to which a theory of consciousness could tie it.

13. 'Dark' here is a highly misleading metaphor. Given the opacity of the skull, all is dark, indeed pitch black, in the brain of all conscious beings, too. Representing light and giving off light are two entirely different things. (Dennett 1988 makes very good use of this distinction.)

14. Exactly what kind of possibility a zombie thought-experiment would have to establish is much debated: logical possibility, natural possibility (possible on some set of laws of nature), nomological possibility (possible on our laws of nature). Fortunately, I do not need to go into this exquisitely scholastic literature because I don't think that zombie thought-experiments can even be coherently thought.

15. I say 'old' because it goes back as far as John Locke (1690).

16. Some commentators suspect that this story is an urban legend but Dennett, in one of the most scientifically literate treatments of pain ever written by a philosopher (1978b, p. 209), gives (now rather dated) references to medical accounts of it.

been assessed for therapy on the basis of differences detectable by MRI in what the researchers call an N400 response to semantically usual and semantically odd sentences ('the pizza was served with piping hot cheese' vs. 'the pizza was served with piping hot socks'). Upon finding a reaction to the latter kind of sentence, therapy was ordered – and patients have eventually recovered enough to walk and talk. Without the reaction, the plug could easily have been pulled on their ventilators (Colin Herein, personal communication, reporting on work in Halifax, NS).

Does the phenomenon of imprisoned minds have any implications for cognitivism about consciousness? To have such implications, I think that one would have to add a further condition, one similar to the condition built into zombie and inverted spectrum thought-experiments and just as unrealistic. Here the required further condition would be that not just behaviourally imprisoned minds but also imprisoned minds whose brains had ceased to function in the relevant ways are possible: not just no difference in N400 response, but no N400 response. Let us call an imprisoned mind that is not making a difference even to the brain a *radically imprisoned mind* (RIM). There could not be the *slightest* reason to think that a RIM existed, of course. However, all anti-cognitivists think they need is the bare possibility, in some sense of 'possibility' (see note 14).

Another, more exotic argument against cognitivism about consciousness flows from externalism about representational content. Externalism is the view, in Hilary Putnam famous (1975) saying, that meaning ain't in the head. The content of representations consists of some relationship between what is in the head and the world. Philosophers who accept this view then go one or the other of two ways about consciousness. Some continue to hold the commonsense view that the element of representations or represented objects of which we are conscious is in the head. They then argue that, since representational content is not in the head, qualia are not representational content. Others hold that if representational content ain't (entirely) in the head, then how something appears (or anything else that the element of representations of which we are conscious consists in) ain't gonna be entirely in the head either (Tye, plato.stanford.edu/entries/qualia, p. 10). The former abandons cognitivism. The latter defends cognitivism – but often at the price of considerable implausibility, depending on which external factor is invoked. (Most of our conscious contents remain exactly the same over even radical change of location of our body, for example, which would seem to rule out direct causal links as an external factor playing a role in conscious content.)

To support the claim that a separation of consciousness and cognition is possible, theorists often appeal to Levine's (1983) explanatory gap. According to Levine, one way to understand the connection between a phenomenon and a mechanism is to understand why, given the mechanism, the phenomenon has to exist. With consciousness, not only do we not know of any mechanism or causal process whose operation has to bring about consciousness, we cannot even imagine what such a mechanism might be like. There is nothing like the same explanatory gap with respect to cognitive functioning, so consciousness is radically unlike cognitive functioning, epistemically at least.[17]

Sometimes such arguments go so far as to conclude that what is distinctive to consciousness is not just not cognitive, it is not even physical. One way of arguing for this is to think of a zombie that is a molecule-for-molecule duplicate of oneself. If a zombie such as this is possible, then the conscious aspect of things is not molecular, i.e., not physical. Another is Jackson's (1986) famous thought experiment concerning Mary, the colourblind colour scientist. Mary knows everything there is to know about the experience of colour, therefore everything *physical* there is to know about the experience of colour, but she has never experienced colour herself. Then her problem is corrected and she experiences colour! Clearly she gains something she did not have before. However, she knew everything physical about colour. Therefore, what she gains must be something nonphysical.

---

17. Levine's explanatory gap is part of what makes the hard problem appear to be so hard, too.

The people we have called system anti-cognitivists agree with other system theorists in viewing consciousness as a property of cognitive systems as a whole, not individual representations, but they have their own arguments hostile to cognitivism. We have already seen an argument of Thomas Nagel's (1974). Nagel argues that the only way to understand what a point of view is, is to have one. The more one tries to adopt a third-person or impersonal point of view on a point of view, the more one moves away from what it is. There is no reason, however, to think that the same restriction to the first-person governs our understanding of representation or other cognitive functioning. If so, there is reason to doubt that consciousness is cognitive.

McGinn (1991) argues for a similar conclusion, that we cannot know how consciousness is linked to the representational activities of the brain, by urging that because our only direct, non-inferential access to conscious states (introspection) is so different from our direct, non-inferential access to brains (perception), we will never be able to find laws bridging the two domains.

Penrose's (1999) argument is indirect, going via computationalism. Relying on Gödel's theorem that a system of arithmetic axioms cannot prove intuitively true propositions that can be generated by the system, he argues that consciousness (and cognition) cannot be computational and therefore must be something weird and wondrous to do with quantum phenomenon. Since the only halfway worked out picture of cognition that we have is the computational one, the conclusion Penrose reaches is so different from any usual cognitive account that we might as well group it with the anti-cognitive approaches.

Searle's (1980 and many subsequent publications) Chinese room is meant to establish the same conclusion that consciousness and cognition are not (just) computational. The direction in which he goes in response is different from Penrose's, however. He argues that consciousness *and* cognition must both be some non-computational biological element, at least in part. Unfortunately, he has never been able to identify a plausible brain phenomenon to play this role but his conclusion is still clearly hostile to cognitive system theory.

Why These Arguments Don't Work

In my view, zombie, inverted spectrum, and imprisoned mind thought-experiments are the most serious challenge to cognitivism about consciousness (together maybe with externalism), but let us take a quick look at some of the other arguments first.

There are familiar, powerful difficulties with each of them. Mary acquires something new when she first experiences colour, to be sure, but there is no reason to think that what she acquires is anything more than a new mode of access to facts she already knows. ('Ah, so that is what light of 640 angstrom units processed through V1 to V4 and integrated in the XYZ cortex is like!') Nagel seems to be just plain wrong. By studying the brain and relating our findings to what people say about their experience (for reflections on this method, see Thompson, this volume), we are rapidly finding out precisely what he says we cannot know: what kind of brain structure a point of view is. McGinn's cognitive closure claims are wildly premature – and again, seem to be seriously threatened by recent work using the method just sketched. Penrose seems to have given little or no thought to the role of heuristics in cognition. Not everything we do cognitively has to be governed by deductively closed rules. And Searle has never responded effectively to the so-called System Response: maybe it is plausible to say that the Chinese room contains no consciousness or content in isolation but if one hooked it up to a visual system and gave it exquisitely detailed control of a body, then the situation is no longer at all clear.

So there does not appear to be a serious threat to cognitivism about consciousness in any of

this. Next, zombies, inverted spectra, and imprisoned minds (specifically, RIMs). The worry we are addressing is that for any cognitive system, we seem to be able to imagine that system behaving as it does, functioning as it does, even being built as it is, without it being conscious or while having conscious contents very different from what we would have in the same situation or while being conscious without that showing in brain or behaviour. A requirement common to all three situations is that the difference in consciousness be compatible with complete similarity in cognition, or cognition and brain, and ensuing behaviour or that the presence and absence of consciousness be compatible with complete similarity in brain and behaviour. This is a crucial requirement because if the difference in consciousness goes with any difference in structure, cognition, or behaviour whatsoever, the desired separation of consciousness and cognition will not have been achieved and no argument that the two are radically distinct will have be given.

Put this way, zombie, inverted-spectrum, and RIM thought-experiments entail (and perhaps rely on) a particularly vigorous form of scepticism about other minds, other conscious minds at any rate.[18] Zombie thought-experiments purport to show that no behaviour, no cognition, even no neural structure is conclusive for the presence of consciousness. Inverted spectrum thought-experiments purport to show that no behaviour, no cognition, even no neural structure is conclusive for the presence of conscious content of a particular kind. RIM thought-experiments purport to show that nothing about brains or behaviour is conclusive for the absence of consciousness. This connection with a deep-dyed and highly suspect form of scepticism should get our suspicions up.

The basis of scepticism about other minds (especially other consciousness) is a lack. W*e do not know what behaviour, cognitive activity, or brain phenomena would constitute or justify an ascription of consciousness.* This is the explanatory gap, one of them anyway. Cross it successfully and the threat of zombie and inverted spectrum thought experiments is removed. RIMs are not quite so easy but the first two first.

Traditionally, philosophy has tried to cross the behaviour/consciousness gap with a single mighty bound, to no great success. Not being a very strong jumper, my approach is to try two bounds instead. The first bound is to justify the ascription of representations. This is easy to do – we have to postulate the possession of representations to explain everything from how people (and other animals) react to the Müller-Lyer arrowhead illusion to children acquiring the capacity to recognize false beliefs in others as they develop. Moreover, nothing in the first bound puts us in conflict with the anti-cognitivist. Zombies and inverted spectrum subjects would have *representations.* It is just that they would not be conscious, or would not have the conscious content that we would have when we had the same representations.

In the second bound we argue that, concerning consciousness, once representations of the right kind are in place, there ain't nothin' left over to be left out. This second move works more straightforwardly with zombies than in inverted spectrum cases, so let us explore it with zombies first. Here is what a philosopher's zombie will be like. It will talk to us about how proud it feels of its kids. It will describe the stabbing pain of arthritis in its hip. It will say how pleasant it is to stretch out on a hammock on a hot summer day. It will report its memories, hopes, dreams, fears just as we do. In short, there will be absolutely nothing in its behaviour, *including its self-reporting behaviour,* that distinguishes it in any way from a normal, conscious human being. Even though, supposedly, it is not conscious, it will *represent itself to itself* as conscious, *feel* pleasure, *have* pain, and so on. And it will do all these things, *ex hypothesi, on the basis of representations and representations of the right kind.*

---

18. Many statements of the problem of knowledge of other minds focus on knowledge that others are conscious. Some extend the sceptical worry to intentionality in representation and behavioural control, too, but, as we will argue shortly, intentionality is not the problem here that consciousness is.

Even the zombie itself will be convinced that it is conscious. So what could possibly be missing? The supposition that it is possible that the zombie might nonetheless not be conscious seems to be a truly perfect example of a supposed distinction that makes – and could make – no possible difference.

Is my argument here verificationist? Don't know. Don't care, either. Whatever you want to call it, the principle here is: for a sentence to be an assertion (intuitively, for a sentence to meaningful), the difference between its being true and being false must mark some difference. The difference need not be detectable – but it must be stateable. In the case of the zombie, we should ask, not just what is but what *could be* missing. Can we so much as form an idea of something that could be missing? If the answer is no, the philosopher's zombie is not fully imaginable, hence not possible, not in any relevant sense of possibility.

Possible objection: 'A thing representing itself to itself in some belief-inducing way is not the same as representing itself to itself in the it-is-like-something-to-have-the-state way.' Response: No? What could this difference consist in? For notice: it would not only be a difference that no one else could detect, it would be a difference that the organism itself could not detect! We shouldn't simply *assume* we can imagine cognition without sneaking consciousness in by the back door. *Show* us that you can do it! As I argued long ago (Brook, 1975), we must at least imagine the relevant difference in enough detail to be able to assert plausibly that if what we imagine came to pass, we would actually have the one without the other. To see the silliness of this 'notion' of a zombie, reflect on a point that Dennett makes: even if we could make some sense of the idea of such a zombie, why on earth should we *care* about what would be missing? It would have nothing to do with consciousness as we encounter it, in oneself or in others (2000, pp. 381-2). (This observation also supports the conclusion reached in Section 4 above.)

The same demolition job can be done on inverted spectrum thought experiments. It must be admitted, however, that some people find this application less persuasive. We can imagine, it is said, that Santa Claus' suit seems blue to Suma when it seems red to me, without anything else changing – without, therefore, this difference showing in her behaviour or in her cognition (inferred from behaviour in any case) or her brain physiology. That means that she will engage in the same additive and subtractive exercises with the colour as it appears to her as we do with the colour as it appears to us. When she combines, for example, a chip of her apparent colour with yellow, she will get orange, not green – and she will say that the combination seems orange to her, not green. She will behave as though the suit appears red to her in all other respects, too. In particular, she will say that the suit seems red to her (she is trained to call apparent blue 'red').

A telling story, again from Dennett. Suppose that a particular shade of blue reminds you of a car in which you had a bad accident and so is a colour to be avoided. Now your colour spectrum is inverted. At first things are fine. The things that used to look blue to you now look yellow and you are not averse to them. In time, however, you adapt to the inversion. (Evidence for this is that you again call 'blue' the shades of colour that others call 'blue', fit these shades into a colour wheel the way others do, and so on.) Suddenly you start avoiding the things that you again call blue *and it is because they remind you of the car in which you had the bad accident.* That is to say, the shade of colour in front of you strikes you as the same as the shade of colour of the car as you remember it. Does the shade of colour in front of you and/or the remembered colour of the car now seem to you as yellow seems to others or as blue does (Dennett 1991, p. 395)?

Here is what I am inclined to say, 'Given that everything, *everything*, is exactly as it would be if the colour now appeared blue to you, what could appearing yellow *consist in* here?" Given that you will say that the colour in question appears blue to you and you will react in every respect as though it does, what could the additional state of affairs of it appearing yellow *consist in?* What difference would correspond to 'X appears yellow to you' being true and it being false?

As we said earlier, many people find the strategy just adopted more persuasive in the case of zombies than in the case of inverted spectra. One reason might be this. We can distinguish a macro from a micro problem of knowledge of other minds. The macro problem is whether a being that behaves like us has a mind or is conscious at all. The micro problem is whether, granting that the macro problem is solved, it has a mind or consciousness like ours. In real life, the macro problem almost never arises. By contract, we face the micro problem a million times a day. ('She says nice things but what does she really think of me?') Thus, while many people are prepared to allow, on analysis, that, well, yes, they can't make sense of the idea of a (philosopher's, i.e., behaviourally, cognitively, and even neurally indistinguishable from us) zombie, a lot of people are by no means as willing to allow this about the idea of a (behaviourally, cognitively, neurally invisible) inverted spectrum. For that reason, it takes more therapy (as Wittgenstein called it) to wean people away from the idea of a (philosopher's) inverted spectrum than from the idea of a (philosopher's) zombie. For example, for inverted spectra of the right kind to be possible, subjects would have to be able to make distinctions that are impossible to make. (For the argument for this claim, see Brook and Raymont, forthcoming, Chapter 3.)

Now RIMs. A way to infer consciousness from behaviour and cognition is not going to help us with whether RIMs are possible, because, *ex hypothesi,* there is no behaviour and cognition. It would be nice to have an argument that showed that RIMs are impossible. But I do not. Instead, I will allow the logical possibility of RIMs and try to show that this possibility is not a problem for cognitivism.

The 'nothing left over to be left out' move does not work for RIMs. We know perfectly well what would be left over when the brain stopped while a RIM continued – exactly the same kind of mind aware of itself in exactly the same kind of way as you and I have. Granted, since the idea of a RIM disconnects conscious minds from everything detectable by others, it leads to some bizarre questions: if RIMs are possible, what argument do we have, for example, that the Esc key on my keyboard could not be conscious?[19] All we have is no evidence that it is conscious – but that is what we would have with RIMs, too. (So can we ever be sure that dear old granny is not still with us, trapped inside the lifeless body that we are about to put in the ground?) Still, bizarre is not the same as incoherent.

Indeed, a more radical form of anti-cognitivism than we have seen so far would be in business. In order for something to be conscious, things must be like something to it, and we have argued that this requires some modes of cognitive functioning. So a RIM would have to continue to have such modes and cognition would have to be just as detachable from the brain as consciousness. We would be pushed to what on our own terms could only be called an anti-cognitivist account of cognition![20]

Curiously, this link is helpful to us. On the RIM story, consciousness and cognition at least still go together. Which means that cognitivism about consciousness would remain an open possibility. Which would be progress.

However, it is progress in the wrong direction. The whole point of wanting to show that consciousness is a form of cognition is to lend support to the project of developing a unified account of the mind and brain. Such an account has to be physicalist – has to argue that both consciousness and cognition are properties a physical system, the brain. RIM stories thus appear to threaten physicalism. But this appearance may be deceiving.

---

19. When Wittgenstein canvasses the idea of someone in pain turning into a stone and asks, "What has a soul, or pain, to do with a stone?" (§283), he may be invoking a similar idea.

20. John Kulvicki pointed this out to us.

Physicalists need not be committed to the view that consciousness *must* be physical. They can accommodate the mere logical possibility not just of RIMs but also ghosts in the machine, ectoplasmic consciousness, and all manner of things that are conscious without being physically realized. How? Physicalists accept the possibility of multiple realizability of conscious and of other cognitive states. If so, physicalists can insist that all *actual* conscious states are realized or implemented in purely physical systems, even that all possible conscious and cognitive states *compatible with the laws of nature* are likewise implemented, and yet allow for the possibility of mental states that are not physically realized.

Before the physicalists among us start to gasp and groan, I hasten to add that this concession doesn't concede very much. The issue before us concerns the physical status of consciousness and cognition. If they do not *have* to be physical, this need be nothing more than an artefact of conscious and cognitive types being anomalous, i.e., not being reducible to physical (in this case, brain-state) types. Since that is also true of clocks and radios and hundreds of other types of things whose types are multiply realizable physically and therefore not reducible to any physical types, the possibility of RIMs does not establish that consciousness or cognition are any more nonphysical than clocks are.[21]

RIM and zombies thought-experiments are asymmetrical. If zombie TEs worked, something could be missing even though everything physical is there. If so, the missing element would have to be nonphysical. But the possibility of a nonphysical realization holds no implications for what something is like *as realized in our world*. Whatever the general reducibility of clock-types, in our world clocks are composed entirely of matter. The possibility of RIMs holds no implications that anything else is true of consciousness or cognition. Consciousness and cognition *could* be nonphysical. So could clocks. But there is no reason to think that any of them are nonphysical, or even that they could be nonphysical in any world with our laws of nature.

Conclusions: Zombie thought experiments cannot be done as specified. Inverted spectrum thought experiments cannot be done as specified. RIM thought-experiments do establish a possibility of consciousness and cognition being nonphysical but not in any way that is not also true of clocks and radios. If anti-cognitivism in its various forms does not succeed, there is no reason to think that consciousness is anything more than an aspect of representing and/or of the cognitive machinery for managing representations. We need not be ignoring consciousness or changing the subject when we look to give a cognitive account of it. Consciousness is safe for neuroscience.

Final note: Whatever the merits of the specific moves made in this paper, I think that it shows how traditional philosophical techniques still have a role to play in neuroscience. These techniques are certainly not neuroscience but they clear away confusions and lay out better ways of thinking about things. If I am really lucky and the particular moves succeed, what would I have accomplished? I would have set things up so that when the next generation puddle about in their hippocampi and thalamuses and V1 and MT, they will know that the cognitive functions and neural implementations of these functions that they find are, or least could be, constituents of *consciousness,* not merely something correlated with consciousness.[22]

---

21. For more on multiple realizability and its implications, see Brook and Stainton 2000, ch. 5, especially the discussion of Descartes' indivisibility argument for mind/brain dualism.

22. Thanks to Kathleen Akins, Dan Dennett, Zoltan Jakab, Jerzy Jarmasz, Luke Jerzykeiwicz, Jamie Kelly, Christine Koggel, Kris Liljefors, James Overall, Don Ross, Sam Scott, Rob Stainton, Edina Torlakovic, Chris Viger, Tal Yarkoni and especially Paul Raymont, who has written a book with me on these topics (forthcoming). This paper is derived from Chapters 1 and 3.

References

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Block, N. 1995. On a confusion about a function of consciousness *Behavioral and Brain Sciences* 18, 227-47.

Brook, A. 1975. Imagination, possibility, and personal identity. *American Philosophical Quarterly* 12, 185-98

Brook, A. and P. Raymont. 2003. Unity of Consciousness. *Stanford Encyclopaedia of Philosophy.* http://plato.stanford.edu.

Brook, A. and P. Raymont. forthcoming. *A Unified Theory of Consciousness.* Cambridge, MA: MIT Press.

Brook, A. and Stainton, R. 2000. *Knowledge and Mind*. Cambridge, MA: MIT Press/A Bradford Book.

Calvin, Wm. 2003. Review of Zeman (2003). *New York Times Book Review* Sept. 28, 2003.

Chalmers, D. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2, 200-19.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P. M. 1995. *The Engine of Reason, the Seat of the Soul.* MIT Press/A Bradford Book.

Churchland, P. M. 2002. Catching consciousness in a recurrent net. In: Andrew Brook and Don Ross, eds. *Daniel Dennett.* New York: Cambridge University Press.

Churchland, P. S. 1983. Consciousness: The transmutation of a concept. *Pacific Philosophical Quarterly* 65 , 80-95.

Crick, F. and Koch, C. 1994. *The Astonishing Hypothesis.* New York: Scribner's.

Dennett, D. 1978a. Toward a cognitive theory of consciousness. In his *Brainstorms.* Montgomery, VT: Bradford Books, pp. 149-73

Dennett, D. 1978b. Why you can't make a computer that feels pain. In his *Brainstorms.* Montgomery, VT: Bradford Books, pp. 190-32.

Dennett, D. 1991. *Consciousness Explained.* Boston: Little, Brown.

Dennett, D. 1998. Real consciousness. In his *Brainchildren.* Cambridge, MA: MIT Press.

Dennett, D. 2000. With a little help from my friends. In Ross, Brook, and Thompson 2000.

Dennett, D. 2001. Are we explaining consciousness yet? *Cognition* 79, 221-37.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Flanagan, O. 1991. *Consciousness Reexamined.* Cambridge, MA: MIT Press.

Flanagan, O. and T. Folger. 1995. Zombies and the function of consciousness. *Journal of Consciousness Studies* 2, 313-21.

Folger, T. 2000. Zombies explained. In Ross, Brook, and Thompson, eds. 2000.

Henein, C. personal communication, Oct. 31, 2003.

Jackendoff, R. 1987. *Consciousness and the Computational Mind.* MIT Press/A Bradford Book.

Jackson, F. 1986. What Mary didn't know. *Journal of Philosophy* 83:5, 291-5.

Levine, J. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly* 654,

354-61.

Locke, J. 1690. *Essay Concerning Human Understanding.* London: Basset.

Mack, A. Inattentional blindness. http://psyche.cs.monash.edu.au/v7/psyche-7-16-mack.htm.

Mack, A. and Rock, I. 1998. *Inattentional Blindness.* Cambridge, MA: MI T Press.

McGinn, C. *1991. The problem of consciousness : essays towards a resolution.* Oxford; USA : Basil Blackwell.

Nagel, T. 1965. Physicalism. *Philosophical Review* 74, 339-56.

Nagel, T. 1974. What it is like to be a bat? *Philosophical Review* 83: 435-50.

Penrose, R. 1999. *The Emperor's New Mind; Concerning Computers, Minds, and the Laws of Physics.* Oxford: Oxford University Press.

Perry J. 1979. The problem of the essential indexical. *Noûs* 13:3-21.

Perry J. 2001. *Knowledge, Possibility, and Consciousness.* Cambridge, MA: MI T Press.

Posner, M. 1994. Attention: the mechanism of consciousness *Proceedings of the National Academy of Science USA*. 91: 7398-7403.

Putnam, H. 1975. The meaning of 'meaning'. *Mind, Language and Reality: Philosophical Papers Vol. 2.* Cambridge: Cambridge University Press, pp. 215-72.

Rosenthal, D. 1991. *The Nature of Mind*. Oxford: Oxford University Press.

Ross, D., Brook, A. and Thompson, D. *Dennett's Philosophy: A Comprehensive Assessment.* Cambridge, MA: MI T Press.

Searle, J. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 417-58.

Shoemaker, S. 1968. Self-reference and self-awareness. *Journal of Philosophy* 65, 555-67.

Tye, M. 1995. *Ten Problems of Consciousness.* Cambridge, MA: MIT Press.

Tye, M. Qualia. plato.stanford.edu/entries/qualia.

Wittgenstein, L. 1953. *Philosophical Investigations,* trans. G. E. M. Anscombe. Oxford: Basil Blackwell.

Zeman, A. 2003. *Consciousness: A User's Guide.* New Haven: Yale University Press.