Chapter 1. A Two-by-Two Matrix[1]

By Andrew Brook

Now is an interesting time to be working on consciousness. From the advent of cognitive science in the 1950s and 1960s up until well into the 1980s, most philosophers and other cognitive researchers pretty much ignored it. The philosopher Daniel Dennett once put the situation this way:

> 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. [1978, p.149]

He could have added that this was pretty much the attitude of most philosophers, too, especially philosophers who viewed themselves as part of cognitive science. In the period in question, one could easily have concluded that in the view of most cognitive researchers, cognitive functioning could proceed perfectly well without any consciousness at all.

This situation began to change in the mid-1980s. The psychologist Bernard Baars (especially1988) developed a methodology that he called contrastive analysis (compare the difference made by performing a task consciously and without consciousness). Now researchers has a technique better than the widely discredited appeal to introspection to study consciousness and consciousness studies began to flourish. Centres of Consciousness Studies sprang up in a number of universities, the most famous being at the University of Arizona. A bit later, an email list called Psyche began and immediately attracted most of the significant researchers on consciousness as members. The Association for the Scientific Study of Consciousness was started: as of the summer of 2001, it had held its fifth international conference. And so on. Consciousness studies is suddenly playing a significant role in cognitive research.

In the years since 1988, two things have happened. There has been a explosion of work. And there has been an explosion of new terminology and new theories – just what one would expect in a field where the subject is an intensely difficult one and serious multidisciplinary work is being done for the very first time. Here is some of the blossoming terminology: access consciousness, phenomenal consciousness, self-consciousness, simple consciousness, creature consciousness, state consciousness, monitoring consciousness, awareness taken to be coextensive with consciousness, awareness distinguished from consciousness, higher order

---

[1] Carleton University Cognitive Science Technical Report 2003-01
URL http://www.carleton.ca/iis/TechReports
© 2002 Andrew Brook

thought, higher order experience, qualia, displaced perception, transparency .... and on and on. Except when discussing something specifically associated with one of these terms, I will for the most part ignore then and develop my own terminology as I need it, connecting it to existing terms when more light than heat would be generated by doing so.

This book will strike many people as eccentric. It does not obsess about 'qualia' and other putative properties of individual psychological states, so it will strike many philosophers as eccentric (though I do say something about qualia). But it does not obsess about attention as the basis of consciousness, either, so it will strike most experimental psychologists as eccentric (though it eventually says a bit about attention). Well, you can't please everyone! Neither qualia as usually understood nor attention captures what is distinctive about consciousness.

Part 1. The Logical Geography of Two Divisions over Consciousness

Reading recent writings on consciousness by philosophers and by cognitive psychologists, one could be forgiven for wondering if they are talking about the same thing. One great divide is over the issue of whether consciousness is a representation or a representational property of some kind. Many philosophers (Jackson, Chalmers, Block, Nagel, McGinn) say no, it is not – consciousness, it being like something to be in a certain state, is different from what a state represents or what makes it a representation. Virtually all psychological investigators (Baars, Posner, Mack and Rock, Shallice, Jackendoff) and some philosophers (Tye, Dretske, Rosenthal, Dennett, P. M. and P. S. Churchland) say that it is – consciousness is attention, or attention plus working memory, or sentience or ... . This division between *representationalists* and what I will call *anti-representationalists* about consciousness is fairly well known.

Another divide has not often been commented on, yet may be even more deep-running and significant. Most philosophers 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. Experimental psychologists by contrast focus on properties of cognitive systems as a whole: global workspace (Baars 1988), intermediate level of processing (Jackendoff 1987), or attention. Attention has been particularly singled out for ... attention. For Posner or Mack and Rock, for example, to be conscious of something simply is to pay attention to it (Posner 1994; Mack and Rock 1998). Let us call this the *system approach* to consciousness. Nor is the latter approach confined to psychologists. Among philosophers, both Dennett (1991) and P. M. Churchland (1995) have built whole-cognitive-system models of consciousness.

These two divisions give us a two-by-two matrix of positions on consciousness:

	Representationalists	Anti-representationalists
Atomists		
System Theorists		

Figure 1: A two-by-two matrix of positions on consciousness

Later in the chapter, I will identify some more specific theorists for each of the four positions. We will also do a preliminary assessment of possible strengths and weaknesses of each. But first we need to explore the two divisions. I will begin with the atomist/system theorist division.

1. Atomist and systems approaches to consciousness

The atomist approach

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, 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 second division – this aspect is or is not something representational – but the important point for us now is that atomists ignore the cognitive system that has states of consciousness. They may add, '... to look red *to me'* but they do nothing with the addition. The nature or role of thing that has such states, the subject of conscious states, plays hardly any role in this literature.

> **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.

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' in Thomas Nagel's (1974) famous phrase.)

Note that atomism as I have just defined it is first of all an epistemological view, a view about how consciousness can be studied – state by individual state. However, it almost always goes with – often by taking for granted – a certain view of what conscious states are like: that such states can be individuated one by one, that they have clear boundaries, that they can have the character they have whatever the character of the states surrounding them and the system of which they are a part. On its face, atomism is a remarkable view. It is not remotely plausible to

think that one could say anything interesting about a conscious state without paying attention to what surrounds it and what the thing for which it is conscious is like.

The system approach

In sharp contrast to atomism, the dominant approach to consciousness in experimental psychology holds that consciousness is a property of the cognitive system as a whole. Let us call it 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.

There is a great diversity of opinion as to what the relevant property is. We cannot begin to explore the whole range of options here but a few examples might be instructive. Baars (1988) holds that consciousness consists in a global workspace of a certain kind. Jackendoff (1987) urges that it is an intermediate level of representation, a phonetic or similar level between acoustic or visual input and full-blown conceptual content. Many theorists link consciousness closely to attention. For example, Mack and Rock: "Attention [is] the process that brings a stimulus to consciousness" (Mack and Rock 1998, xx), "if a ... percept captures attention, it then becomes an explicit percept, that is, a conscious percept" (Mack, 2001, 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, 7398]

Nor is this approach without its philosophical allies. Dennett's (1991) multiple drafts model is one 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, 74]

The system approach to consciousness once dominated philosophy – think of Descartes and Kant. System approaches are not a monolith. One division is particularly important. Some system theorists about *consciousness* are atomists about *representation* and some are not. For

system theorists who are atomists about representation, even though consciousness is a property of representing systems, not individual representations, representations themselves come in discrete units that can be individuated one by one, have clear boundaries, and can have the character they have whatever the character of the states surrounding them or the system of which they are a part. To summarize,

> **Atomist view of representation –** representations are discrete units with clear boundaries separating one from another

The view that to be conscious of something is to pay attention to it is a nice example; attention is a property of whole cognitive systems but, many hold, one can pay attention and thereby become conscious of individual representation (or what individual representations are representing) one at a time. Many theorists who explore consciousness experimentally accept this combination of positions: they take a system approach to consciousness but are atomists about representation.

Other system theorists about consciousness also take a system approach to representations. They deny that representations come in neatly individuated, discrete units that have clear boundaries, etc., etc. For them, representing too is a property of whole cognitive systems.

> **System view of representation –** there is no such thing as discrete individual representations with clear boundaries separating one from another

On this approach, representing may be viewed as a matter of certain patterns existing in the behaviour of a whole cognitive system (Dennett),[2] as multidimensional phase-space processing of information widely distributed across the brain (P. M. Churchland), or even as a result of intersubjective triangulation (Davidson), but never as an assembly of small, discrete units. These theorists certainly think that the cognitive system does lots of representing – but not via anything remotely resembling discrete individual representations as conceived of in the tradition. (Note: if atomism about representation is false, atomism about consciousness will almost certainly be false, too. We will return to this implication in Section xx.)

2. Is Consciousness Something Representational?

Next the division over whether consciousness is a kind of representation or a representational property of representations (and maybe other, nonrepresentational states). To isolate what is at stake here, we need to make a couple of distinctions. The first is a distinction between

---

[2.] This is where I think Dennett fits. In Dennett's view (see 1991b, for example), these representational patterns are perfectly real, but there are no discrete, separable representations behind them. (Since I want to keep the early sections of this introductory chapter relatively uncluttered, I will try to confine comments on the work of others footnotes as I have done here.)

consciousness in connection with what might not itself be a representational state (feeling good or a mystical sense of oneness with the universe) and consciousness in connection with what clearly is a representational state, how something looks, for example. Set the former cases aside (we will return to them in Chapter 5). Even when a representation is clearly involved, is *the consciousness* representational? The second distinction we need to make is between conscious and nonconscious representations. Probably the vast bulk of our representations never make it to consciousness, certainly not consciousness *of* the representation, and that includes some representations that are cognitively very active. But the question now before us is, even when we *are* dealing with consciousness in connection with a representation, can the difference between conscious and nonconscious representation *be captured* by appealing to representational properties of any kind? Here are the alternatives.

> **Representationalism** – the view that the difference between a state being and not being conscious is a difference in how that state represents or a difference in the kind of representation it is or a difference in something representational.

And

> **Anti-representationalism** – the view that the difference between a state being and not being conscious is not a difference in how that state represents anything or a difference in the kind of representation it is or a difference in anything else representational.

Representationalism, specifically atomist representationalism, first.

There are (at least) three current forms of atomism about consciousness that embrace representationalism.

One form holds that in the same way as things can appear to me in my experiences, my experiences can appear to me; appearing to me as a perception of red is a property of my experience of red. This view goes back at least to Kant and probably a good long way before that. We have conscious representations of the world, and we have conscious representations of ourselves and some of our states (Lycan 1987, 1996 – check; Flanagan 1992; Tye 1995). On the one hand, being conscious of ourselves and our representations is more than representing something in the world. We know more things and different things about ourselves and our representations than we could gain by any process of inference (see the second form immediately below). On the other hand, far from the "felt quality" and the functional properties of representations being separable, in all such cases we are dealing with one and the same thing. What makes the appropriate representations conscious? Opinions vary but one common answer is that we can describe them, report on them (Lycan 1987, 1996).

Another form holds that while we are conscious *via* conscious states, we are not conscious *of* conscious states. We are directly conscious only of what is presented in such states and consciousness of oneself and one's states is an inference, namely, the inference from our experiencing something in the world to the fact that we are experiencing it. When I experience red, I can infer that I am having an experience of it. Being painful is not a property of an experience of pain, it is a matter of experiencing something painful. These theorists urge that

conscious states are "transparent": to the extent that our conscious states appear to us at all, they do not appear as anything different from the states of affairs of which they make us conscious (Dretske 1995). Hence it has come to be called transparency theory.[3]

In the third form, a representation of red gets to be conscious by becoming the object of a thought or other representation. This is the higher-order thought (HOT) and higher-order experience (HOE) approach. The term 'higher-order' refers to the fact that the thought is about another psychological state, not some state of affairs in the world (Rosenthal 19xx; others xx).

For an anti-representationalist, by contrast, consciousness is quite unlike anything else known to exist in cognitive systems. In particular, consciousness is not a representational property of anything, not even when the representation in question is a conscious state. Indeed, consciousness does not, or need not, even causally interact with other representations and behaviour (Nagel 1965, Block 1995, Chalmers 1996).

Note that representationalists and anti-representationalists need not disagree concerning the nature of representations. There are atomists about representation – and consciousness – who are representationalists (Tye 1995, Lycan 1987) and who are anti-representationalists (Block 1995, Chalmers 1996). Equally, among system theorists who are both atomists about representation and system theorists about consciousness, and system theorists who take a systems view about both representation and consciousness one can find representationalists (Posner and Mack and Rock; Dennett and P. M. Churchland) and anti-representationists (Nagel, McGinn, Searle and Penrose – system anti-representationalists all while holding views on the nature of representation that are all over the map of possibilities).

Here is how anti-representationalist can get going within atomism about consciousness. When something appears to us to be a certain way, the representation in which it appears can play two roles in our cognitive economy. On the one hand, 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

[3] The difference between the first two varieties of atomism can be brought out by reference to consciousness of own's own experiences. Often when we are conscious of something, we are representing that thing. For example, when we are conscious of something in the world around us or, in proprioception, a state of our own body, we are representing that thing. But what about the case where we are conscious of the experience, the representation, itself? Here the literature splits. Many hold that when one is conscious of one's representations, one is representing the representation, i.e, something different from what the representation itself represents. This yields the first kind of atomism. However, some theorists argue that consciousness of one's representations is also given merely by representing something in them. Focus on a square red patch. Now focus on your *awareness* of that square red patch. Are you now aware of new properties? Some philosophers answer 'no' (Harman (1990), Dretske (1995) Tye (1995) and Crane 19xx). On this so-called transparency theory of consciousness, consciousness of one's own experience is merely a displaced perception of something in the world (Dretske 1995, Tye 1995), i.e., something that we infer from what representations do make us directly aware of, namely, states and events in the world, not something represented to us directly. Transparency theory is a central topic of Chapter yy.

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 case, 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. My representation could do these jobs for me just as well even if I were not aware either of the stick or of my representation of it. But I am also *conscious* of the stick – it does *appear* to me in a certain way. This can easily seem to be something different from any representational properties of the representation, at any rate properties such as those we just considered.[4]

Arguments for this "neo-dualist" conclusion, as Perry (2001) calls it, often rely on thought experiments about inverted spectra (how colours appear to us could be inverted without changing how our representations of colour function as representations), zombies (there could be creatures for whom it is not like anything to represent whose representations nevertheless function cognitively just as representations function in us), and appeals to externalism about mental content and colour-blind colour scientists (to both of which we will return in a moment). All such thought experiments are designed to show that the content and causally effective aspects of a representation, on the one hand, and how it appears, on the other, could come apart. Thought experiments about inverted spectra and zombies are particularly prominent.

To support the claim that this separation 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 to understand why, given the mechanism, the phenomenon has exist. With consciousness, not only do we not know of any mechanism or causal process whose operation has 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.[5]

Another, more exotic argument against representational theories of consciousness flows

---

[4.] 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 problelm, 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-representationalism, the two phenomena will at least be quite *different* problems. One aim of this book is to put some pressure on that idea.

[5.] Levine's explanatory gap is part of what makes the hard problem appear so hard, too.

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 qualia, the element of representations of which we are conscious, are 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 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 to be in the head either (Tye, plato.stanford.edu/entries/qualia, p. 10). We will explore the implications of externalism for representational theories of consciousness in Chapter yy.

Sometimes such arguments go so far as to conclude that what is distinctive to consciousness is not just not representational, 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 'qualia' (the conscious aspect of things) are not physical properties. 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.

Among system theorists, representationalism is by far the dominant view. All experimental system theorists (we have seen the examples of Posner and Mack and Rock) and a good number of philosopher system theorists (Dennett, the Churchlands) are representationalists. Representationalist system theorists share the representational atomist's view that consciousness is just an aspect of representing of some kind. The difference between them is that, while the atomists takes consciousness to be something that can be studied in individual representations, system theorists take it to be a property of whole cognitive systems of some kind.

Quite a number of philosopher system theorists are anti-representationalists, however. Here are some examples. Consciousness is a biological property of brains (Searle[6]), Consciousness is a quantum phenomenon of some kind (Penrose). We can never know how consciousness is linked to the representational activities of the brain (McGinn). An argument of Thomas Nagel's (in 1974) is particularly interesting. Nagel argues that to understand what a point of view is, the only way is to have one. The more one tries to grasp what a point of view is

6. Searle's relationship to anti-representationalism is tricky. On the notion of 'representation' we have been using here, he is anti-representationalist. But he might be anti-representational about representations so understood, too, depending how closely we want to link representing to formal information processing operations. For Searle, representing is a matter of biology, consciousness included.

by considering it from an third-person or impersonal point of view, the more one moves away from grasping what it is. But science, the study of the physical, is done from an impersonal point of view. If so, what a point of view is like will elude science. There is no reason, however, to think that the same thing is true of representation. There is no reason, therefore, to think that consciousness is representational.

Whether any of these thought experiments establish anything is a real question. Indeed, it will be the main question of Chapters 3 and 4. Against anti-representationalism in all its forms, representationalists maintain that consciousness is simply an aspect of representing. Far from consciousness being a "layer" that could be peeled off a representation or representational system while leaving everything cognitive and behavioural intact, as in the zombie story, consciousness *simply is* representation: some kind of representation or some aspect of representing.

Let us pull our results so far together. Recall Figure 1:

	Representationalists	Anti-representationalists
Atomists		
System Theorists		

Figure 1: A two-by-two matrix of positions on consciousness

We can now put some names in the cells.

	Representationalists	*Anti-representationalists*
Atomists	Lycan, Tye, Rosenthal, Dretske	Chalmers, Block, Jackson, McGinn (?)
System Theorists	Posner, Jackendoff, Dennett, Paul Churchland	Nagel, Searle, Penrose, McGinn (?)

To be sure, there are important disagreements inside some of these cells. While Rosenthal and Dretske both approach consciousness atomistically and representationally, for example, Rosenthal thinks that consciousness is a matter of higher-order thoughts while Dretske denies this.[7]

On the anti-representational side, Chalmers thinks and Jackson thinks or thought that

[7.] Separately from this disagreement, Rosenthal also sometimes talks as though he wants to be in the systems camp (see 2002, for example). Something fairly big may be hiding here, so we will return to the issue when we examine his HOT model in Chapter 4.

consciousness is something nonphysical, while McGinn (1991) holds it to be something physical. Chalmers thinks that we may eventually understand consciousness, nonrepresentational though it is, while McGinn denies this. In turn, there are different ways of denying that we will ever understand consciousness. We have seen Jackson's (1986) argument that something about what we learn in direct experience cannot be captured in any theory about physical systems, and Nagel's (1974, 19xx). argument that the nature of a point of view is systematically inexpressible in an objective theory done from an impersonal or God's-eye point of view. Drawing on Chomsky's notion of cognitive closure, McGinn comes at the conclusion another way. He argues that the nature of consciousness will forever remain as closed to us as physics is to a field mouse. Pinker (199x) holds a similar view. Flanagan (1992) calls these people collectively the new Mysterians.

There are also big disagreements inside the system representational camp. For example, Churchland and Dennett are both card-carrying members of this group, yet Churchland thinks that consciousness is a biological property of cognitive systems, Dennett thinks that it is induced by culture.

The question-marks after McGinn's name indicate that where to place him is not straightforward. Is his position a system position or an atomist one? He claims that we cannot know what consciousness is like because we have no way of knowing what the laws bridging conscious states to brain states are. Depending on what kind of laws he has in mind, he may be an atomist or a system theorist. If he focuses on laws bridging individual kinds of conscious states to individual kinds of brain states, then he is probably an atomist. If he focuses on bridging consciousness as a system to the brain as a system, then he is a system theorist.

Nonetheless, our two-by-two matrix does capture two major differences in contemporary work on consciousness and nicely organizes many of the key players. But how nicely?

3. Eliminativism: A Third Division?

Do the two divisions we have identified exhaust the currently active alternatives on consciousness? It may appear that there is a third, between eliminativists and non-eliminativists about consciousness..

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

How could 'consciousness' turn out not to be a theoretically useful term? That would be the case if the term is merely a misleading name for the various processes much better named and described by other terms, if there is nothing in us very much like the picture built into our term 'consciousness' pretheoretically. Perhaps the term has been used for such a wide and diverse group of phenomena that it should be replaced with a number of more specific terms (P. S. Churchland 1983). Or perhaps when we finally understand how the brain processes information,

it will turn out that there is nothing there that it would be theoretically useful to label ‘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.)

Contrary to what is often thought, right now there are no clear eliminativists about consciousness. At one time, Patricia 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 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.

None of this is surprising. Unlike the notion of intentionality, the notion of consciousness is not a term of art. It is a notion that 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 are describing nothing real, or even a bunch of things better discussed in a different vocabulary (thought doubtless some of the latter will turn out to be the case).

Some think of Dennett’s (1991) multiple drafts model of consciousness as eliminativist. This would be quite wrong. Dennett certainly rejects a dominant way of thinking about consciousness, which he calls Cartesian materialism. But this is to reject a *theory* of consciousness, 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 stable, less universal 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 just want to deflate philosophical pretensions about what consciousness is like (2000, pp. 369-70). (How he can be a realist of this sort about consciousness given other things he says is another question, a question to which we will turn at the end of Chapter 4.)

Seeing that there are no genuine eliminativists about consciousness currently about reveals something about the nature of the debates introduced discussed in the previous sections. Anti-representationalist use zombie thought-experiments and such to argue that system theorists and representationalists in general end up ignoring what they are trying to explain, namely, consciousness. When they do so, however, they are not accusing their opponents of being eliminativists. They do not charge that their opponents *intend* or even *want* to leave out consciousness. To the contrary, all parties to the debates we have been examining take consciousness to be a perfectly real phenomenon in need of explanation and they want their theories to explain it, not eliminate it. The charge against representationalists is that they fail to do so.

4. Realism and Consciousness

Some will think that we have overlooked another important division, a division between those who take consciousness to be a real phenomenon, as real as information processing in the brain and bread and circuses in the world (realists) and those who qualify or deny this claim (irrealists). (Note: irrealism encompasses all 'deviant' positions on the reality of consciousness, not just flat-out anti-realist ones.)

The theorists of consciousness introduced above have positions on the reality of consciousness, sometimes diverging positions. Atomists, for example, are all or nearly all local realists (we will introduce this term in a moment), system theorists mostly are not. However, there is also real convergence on the issue.

Concerning realism, the interesting positions are the position that the physical world is real and consciousness is just as real, on the one hand, and the position that the physical world is real but consciousness has a different reality status of some kind, on the other. A form of irrealism that swallowed everything, or even all of science, would not be about any difference with respect to reality between consciousness and anything else. (Kant's distinction between transcendentalism idealism and empirical realism is casting a shadow here.)

Dennett and Davidson are two philosophers who will immediately come to mind when the question of the reality of consciousness is raised. Dennett we just discussed so let us focus on Davidson. Davidson is clearly a system representationalist about consciousness, though he has a very special, system view of representation, too. How does he view the reality of consciousness?

One way to deny that consciousness is real would be to argue that, while we are doing *something* useful when we talk about consciousness, what we are doing is not referring to a state of anything, not a representational state, not an attentional state, not even a real pattern in the system's behaviour (Dennett). Rather, attributing consciousness to something is a way of interpreting it, a useful (or even essential) means to explaining and predicting behaviour but nothing more than that. Is this Davidson's (19xx) view? For Davidson, consciousness arises out of a complex triangular interaction among oneself, other purposive beings, and the world. By itself, this triangulation picture could be fully realism; the result of the triangulation, consciousness, could still be a real property of cognitive systems. For Davidson, however, not only does consciousness arise out of triangulation, it is (roughly) nothing more than triangulation. When triangulation results in stable attributions of consciousness to self and others, that is what consciousness is. Moreover, if one's notion of what is going on in any part of the triangular pattern changes, one's notion of what is going on everywhere else in the pattern will probably change, too – the consciousness that one attributes is what fits best overall. Finally, one's assessment of the triangle is governed by norms. The triangular pattern must approximately satisfy certain norms, norms that are 'constitutive principles' of discourse of this

kind. Is all this compatible with taking consciousness to be a real, detectable property of anything?

The answer to this question is a matter of some controversy but I am inclined to say that the correct answer is, 'yes'. (I am inclined to think that this is Davidson's view, too, but that is a story for another day.) The property in question will be a rather smeared out one, ranging over the whole triangle but, once it has been established that the norms are appropriately met, it is not clear to me why a real property of *some kind* has not thereby been revealed. Certainly the property, consciousness, will be radically unlike what atomist anti-representationalist take consciousness to be like, but for all that consciousness in Davidson's vision would still be a real property.

What the previous paragraph may indicate more strongly than anything else is simply that the notion of the 'real' is not a very clear one here. Let us hope that we do not have to clarify it. Aside, perhaps, from Rorty and people influenced by him, virtually everyone writing about consciousness considers it a real property, a property there to be detected in the right kinds of critters by the right kinds of critters (where, more centrally in this case than any other, the detector and the detectee may be the same critter). What they disagree about is what consciousness really is, not whether it is something real. If so, we can devote our time and attention to issues about what it is and say no more about whether it is. (Actually, as I just said, we will say a bit more about realism in Chapter 4 in connection with Dennett.)

To be sure, the various positions organized along these two dimensions will have important differences about *what* is real when consciousness is present. In particular, atomism about consciousness goes with what I called *local realism* a short time ago. Local realism is the view that what makes consciousness real is that it is a real property of individual psychological states (or, as in HOT and HOE theories, a very small groups of such states).

> **Local realism** – the view that consciousness is a real property of individual psychological states or of small numbers of psychological states.

Local realism, then, views consciousness as either a (presumably nonrelational) property of single psychological states or a relational property but one that ties only very small groups of psychological states to one another.

This notion of local realism is not very precise but it is precise enough. What matters is the contrast. Local realism contrasts with two views. One is irrealism about consciousness, which we just discussed. The other is what we might call *global realism*, the view that consciousness, though a real property, is not a property of individual psychological states or small groups of them. Rather, it is a property of a whole cognitive system.

> **Global realism** – the view that consciousness is a real property of a whole cognitive system.

What is important is that one can reject the view of representation that underlies local realism without rejecting realism. I mention this here because most system theorists, especially system theorists who study consciousness experimentally, are realists about consciousness; but they are

not *local* realists.

Does atomism require local realism? Require? It is hard to tell. Most atomists *are* local realists and it would be hard to see why a theorist about consciousness would think that she could isolate individual psychological states or tiny groups of such states for study if s/he did not hold that consciousness is there to be found in them. Anyway, this question gets interesting only if we have some reason to take atomism serious, a topic that we will take up in Chapter 4.

Part 2. Problems and Requirements

5. Importantly different system theories

Let us now turn to some of the problems facing the various approaches to consciousness and a few strengths. The problems arise from a number of very different directions:

- background issues about the nature of representation
- differential implications of two varieties of consciousness
- requirements on a theory of consciousness
- the apparent separability, at least conceptual separability, of consciousness and cognition

Background issues about the nature of representation first. As we saw, system representationalism comes in very different versions. There are theorists such as Posner and Jackendoff who, while realists and system theorists about consciousness, are atomists about representation. Then there are theorists such as Dennett and Churchland who are system theorists about both representation and consciousness. For these latter theorists, while representing is a perfectly real feature of cognitive systems, there are no representations of the sort envisaged by atomists about representation. Nothing representational comes in discrete units with clear boundaries separating one from another

This view of representing has immediate implications for atomism about consciousness. If atomism about representations is false, atomism about consciousness has to be false.

> **Problem for the atomist approach –** If atomism about representations is false, then atomism about consciousness will also be false.

6. Two kinds of consciousness

The variety of different things that we can have in mind when we use the word ‘consciousness’ is a big topic. We will see just how big in the next chapter. Moreover, there is no agreement on the most basic issues in connection with this usage. Some theorists use ‘consciousness’ to refer both to a kind of access to the world (sometimes called ‘consciousness of the world’ or even

'simple consciousness') and a kind of access to oneself and one's own states (sometimes called 'consciousness of self' or 'self-consciousness'). Others take the term to refer only a kind of access to onself and one's own psychological states. (The latter is roughly the way psychoanalysts use the term: unconscious states are states of oneself to which one does not have the requisite kind of access.) This confusion of uses of the term is endemic and extremely unfortunate.

We will attempt to achieve some clarity about the matter in the next chapter. For now, let me introduce one distinction. If it has merit, it reveals something quite interesting about anti-representationalism. The distinction I have in mind is between,

> **Consciousness of the world –** the kind of consciousness that is present when we are conscious of the world around us,

and,

> **Consciousness of self –** the consciousness that is present when, for example, we are conscious of *representing* items in the world[8]

If this distinction (or something roughly extensionally equivalent to it; see note 10) is sound, a problem for anti-representationalism appears. Anti-representational accounts seem to have nothing to say about consciousness of the world. Anti-representationalism is about the felt quality of psychological states (in its atomist version) and, for example, what a point of view is (in Nagel's system version). Take the problem of inverted spectra, the idea that how something looks to me might be inverted relative to how it is or how it looks to others. Without consciousness of how something looks and limited to consciousness of the world, this would have to come out: how the thing is represented as being and how it appears might be inverted.

---

8. Strictly speaking, the distinction is not between consciousness of self *and* consciousness of the world, because consciousness of self *is* consciousness of an item in the word. As we will see in Chapter 5, a better way to capture the distinction is to distinguish between states of affairs of which we become conscious by perception (and maybe in some other ways) and states of affairs that we become conscious of by having them, feeling them, or doing them. The latter is the way one becomes conscious of one's own perceptions, feeling and sensations, and actions. Consciousness of oneself as subject is in yet a third case. It is closely related to the second but it is quite different from consciousness of a being who is in fact oneself via perception and so on. (This is the consciousness one has when one sees a person who is in fact oneself in a distant mirror or a video and does not know that it is oneself.) Noting that this more complex account is more accurate, we can continue to use the distinction between consciousness of self and consciousness of the world in the current discussion. (Thanks to Edina Torlakovic for insisting that making the distinction in the more straightforward way is not enough.)

A full account of kinds of consciousness divided along these lines, moreover, would have to distinguish between consciousness of self and consciousness of one's own representational states, and between consciousness of real things and events of all kinds and consciousness of fictitious and imaginary objects. Fortunately, we do not need these further distinctions for current purposes but they will become important in later chapters.

But what could this mean? How it appears simply *is* how it is represented as being.[9] When we are dealing with consciousness of our own states, there is at least room for an question to arise whether how things appear to me in that state is how they are being represented as being, how they are being represented as judged by the effects the representation has on other elements in cognition. When one is not conscious of the representing, only what is represented, there is no such room. A problem for anti-representationalism.

Not all forms of representationalism are home and dry on this issue, either. For HOT theories, for example, all consciousness, at any rate all HOT consciousness, has to *be* consciousness of self: for a psychological state to be conscious is for it to be the object of another psychological state.[10] To pull the two problems together:

> **Problem for anti-representationalism and HOT forms of representationalism –** Anti-representationalism and HOT theories have nothing to say about consciousness of the world.

Some other forms of representationalism, by contrast, have something to say about both kinds of consciousness. In transparency theory, for example, the only consciousness there is is consciousness of the world. Attention-based forms of system theory not only have something to say about consciousness of the world, they generally focus on it: when theorists who view consciousness as closely related to attention talk about paying attention to something, for example, they generally have in mind paying attention to something in the world, not paying attention to one's own states.

One response to the suggestion that these two problems exist, of course, would be to urge that what I am calling consciousness of the world is not a form of consciousness, at least not by itself. We will take up this issue in the next chapter.

7. What a theory of consciousness should be able to explain

When we turn to what we want a theory of consciousness to be able to explain, other problems appear. 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 . . .

---

9. *Mutatis mutandis*, the same problem would appear for zombie-ism here. This would be the idea that one could represent the world without the world appearing to be like anything. Not a promising idea!

10. If this view could be sustained, we would have a motive for making a sharp distinction between creature consciousness and state consciousness as Rosenthal and others do.

- reappears in dreams.[11]

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

- Consciousness of oneself and consciousness of one's particular representations, desires, 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 (197x) 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. (Many forms of atomism fall short as early as 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 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 to consciousness.[12]
- Usually a cognitive system is conscious of whole groups of representations in one 'act of consciousness'.
- Usually when a cognitive system is conscious of whole groups of representations, it is also conscious of itself as the common subject of these representations.

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

[12.] 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. Some kinds of conscious states may not be representations, mood states for example, or mystical states. (I think they are, but I will argue that at the beginning of Chapter xx, not here.)

Explaining the features we find on these two lists and perhaps others is a central demand on a theory of consciousness (and a central task of this book). Yet when faced with a list of features of consciousness such as this, the atomistic approach to and all anti-representational models of consciousness just claws the air – and it is hard to see how it could do anything else.

> **Second problem for the atomist approach and second problem for anti-representationalism –** Atomism and anti-representationalism cannot explain many important features of consciousness.

8. Is consciousness any form of cognition?

The suggestion is that atomist and anti-representation approaches to consciousness face major problems. But representationalism faces a problem, too, representationalism of both the atomistic and the system varieties. Indeed, most people in the anti camp simply find representationalism frustrating; for them, it is not talking about *consciousness.* It is either just changing the subject and talking about something else or if it is talking about consciousness, it is missing the most interesting and central features of it. A passage from Dennett illustrates how the frustration arises: "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). Dennett then notes that Baars himself had said something similar in 1999. Statements like this make anti-representationalists want to tear their hair out!

Their sense of frustration is not helped by the tendency of many representationalists to just ignore anti-representationalism and the argument advanced in its favour. At a recent conference at Carleton University, a conference who main papers will soon appear in a book on philosophy and neuroscience (Brook and Akins, forthcoming), there were five papers on consciousness. All of them took a systems representationalist approach and not one of them mentioned that representationalism is rejected in its entirety by many philosophers.

Dennett himself has been fighting in the consciousness wars for far too long to neglect the opposition in this way himself. Indeed, in the very paper just cited he says that he will "diagnose some instances of backsliding and suggest therapeutic countermeasures." We will examine, indeed greatly enlarge upon, some of those countermeasures in Chapters 3 and 4. The point here, however, is that many philosophers would insist that whatever a global workspace model is talking about, it is *not* talking about *consciousness!* Why? Because it is perfectly easy to imagine a global workspace grinding away doing its thing with no consciousness at all.

This worry can be generalized. For any form of representation and any representing system that one could imagine, couldn't such a system do all the wonderful cognitive things that it does without consciousness? This worry even cuts across the system/atomistic divide. It is an issue for all forms of representationalism.

Problem for representationalism – Any representation and any representational system could do the cognitive things that they do without consciousness.

The best known expression of this worry is the zombie thought-experiments introduced earlier: couldn't there be creatures just like us behaviourally, cognitively, or even physically who nevertheless are not conscious?[13] Though they are built and behave in ways wondrously like us, all is 'dark' inside.[14] What if anything zombie thought-experiments establish is hotly contested but if they establish anything, it would seem that all forms of representationalism are in trouble.

Here I want to make just two points about the frustration anti-representationalists feel with attempts to mount representational theories of consciousness. First, whatever sympathy I may have shown in the paragraphs above, the onus *is* on them. Both camps agree that most conscious states are representations (free-floating anxiety, mystical states, and so on are the putative exceptions). Since the anti-representationalists argue that consciousness is also something more, they have to make the case. Absent some reason to believe that consciousness is something more than representations, the rational thing to believe is that it is just representations. (If we can believe in one thing for whose existence we have no evidence, there is nothing to stop us from believing in ten – or ten thousand. This is just Occam's Razor.) Second, however, we cannot just ignore anti-representationalism. If one wants to argue that consciousness is just a kind of representation (and this one wants to argue just that), one must *show that* the attempts to demonstrate that it is something more do not work. One must show that there is nothing to the zombie, inverted spectrum and related thought-experiments. One cannot, as I said, just ignore the issue.

So if we think, as I do, that the place to look for an remotely adequate model of consciousness is system representationalism, we should take on two tasks. The first is a negative one. We should assess the merits of views hostile to the undertaking, atomism and anti-representationalism. Only then, the ground cleared of sceptics and detractors, will people be able to take a genuinely open-minded look at our own model, free of worries about whether what it is talking about is really consciousness. Absent a really good argument that there is no viable alternative to some form of system representationalism, a hint of simply not being on topic is

---

13. 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 colourblind colour scientist thought-experiments) raise similar questions but here we will confine ourselves to zombie thought-experiments. See also Polger's (2000) followup article and Dennett's reply. Note that zombie 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.

14. '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 in another context, as we will see in Chapter xx.)

going to cling to all such models.

9. The Strategy

Here in detail is what we are going to try to do. In Chapter 2, we will try to get clearer about what we are talking about when we use the word 'consciousness' and what a successful model of consciousness has to contain. Then, starting from some moves made influential by Daniel Dennett, we will attack anti-representationalism (Chapter 3) and atomism (Chapter 4). At the end of Chapter 4, we will return briefly to the issue of realism. Chapter 5 begins the positive story, the search for a good system representationalism theory. We start by examining an unnecessary and problematic picture of representation assumed by many forms of representationalism and, via a critique of higher-order thought (HOT) theories, try to build a better picture. Since it will have turned out that what HOT theories are really after is unified consciousness, we will next take up the unity of consciousness (Chapter 6), both synchronic (unity at a given moment) and diachronic (unity over time). Another topic flowing directly from Chapter 5 is consciousness of self. It will be the topic of Chapter 7. With this we will have the skeleton of a system representationalist account in place. We will then turn, in Chapter 8, to a major challenge to all representationalist theories of consciousness, namely, externalism about mental content, examining among other things the implausible transparency thesis about consciousness. (Externalism, we said, is the idea that the content of my mental states is a relational property made up of states of me being in a certain relationship to states of the world, not a property of my states by themselves. One problem, recall, is that my consciousness seems to be entirely a property of me, so externalism puts pressure on all forms of the idea that consciousness is a property of representations.) Finally, in Chapter 9 we will try to put some flesh on our model's bones.

In short, the book has two major targets: anti-representationalism and atomism about consciousness. And it has one major goal: to lay out a system representational model of consciousness that is superior to existing ones. Here is the list of problems that we have uncovered so far.

Problem for the atomist approach – If atomism about representations is false, atomism about consciousness will also be false.

Problem for anti-representationalism and HOT forms of representationalism – Anti-representationalism and HOT theories have nothing to say about consciousness of the world.

Second problem for the atomist approach and second problem for anti-representationalism – Atomism and anti-representationalism cannot explain many important features of consciousness.

Problem for representationalism – Any representation and any representational system could do the cognitive things that they do without consciousness.

More problems will appear in the next chapter.

Thanks: Luke, Don, Tal, Louise, Edina, James, Kris

References [incomplete]

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

Brook, A. 2000. Unity of Consciousness. *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu

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

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

Davidson, D. 19xx. Subjectivity, objectivity, intersubjectivity. *Journal of Idealist Studies* xx

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

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

Dretske, F. 1995. *Naturalizing the Mind*. MIT Press

Jackendoff, R. 1987. *Consciousness and the Computational Mind* MIT Press

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

Lycan, Wm. 1987. *Consciousness* Cambridge, MA: MIT Press

Mack, A. 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

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

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

Rorty, R.

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

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

10,150 words