Delivered version 28.12.2006 in NY

REPLY TO PROFESSOR DENNETT AND PROFESSOR SEARLE

M.R. Bennett and P.M.S. Hacker

Conceptual elucidation

In our book, we aimed to contribute to neuroscientific research in the only way that philosophy can assist science – not by offering scientists empirical theories in place of their own, but by clarifying the conceptual structures they invoke. One of us has spent his life constructing empirical theories about neuronal functions. But those endeavours, which deal with the foundations of neuroscience, provide no part of its *conceptual* foundations. The systematic elucidations we gave of sensation, perception, knowledge, memory, thought, imagination, emotion, consciousness and self-consciousness are not theories. Their purpose is to clarify the psychological concepts that cognitive neuroscientists use in *their* empirical theories. The conceptual clarifications we gave demonstrate numerous incoherences in current neuroscientific theorizing. They show why the mistakes are committed and how to avoid them. Nothing we said debars neuroscientists from introducing novel concepts, but our concern was with the conceptual apparatus they *currently* use.[i]

Cognitive neuroscience is an experimental investigation that aims to discover truths concerning the neural foundations of human faculties. A precondition of truth is sense. If a form of words makes no sense, then it won't express a truth. If it does not express a truth, then it can't explain anything. Philosophical investigation into the conceptual foundations of neuroscience aims to disclose and clarify conceptual truths that are presupposed by, and are conditions of the sense of, cogent descriptions of cognitive neuroscientific discoveries and theories.[ii] If conducted correctly, it will illuminate neuroscientific experiments and their description, and the inferences that can be drawn from them.

Two paradigms – Aristotle and Descartes

Philosophical reflection on human nature, on the body and soul, goes back to the dawn of philosophy. The polarities between which it fluctuates were set out by Plato and Aristotle. According to Plato,

and the Platonic-Christian tradition of Augustine, the human being is not a unified substance, but a combination of two distinct substances, a mortal body and an immortal soul. According to Aristotle, a human being *is* a unified substance, the soul (*psuch_)* being the form of the body. To describe that form is to describe the characteristic powers of human beings, in particular the distinctive powers of intellect and will that characterize the rational *psuch_* . Modern debate on this theme commences with the heir to the Platonic-Augustinian tradition, namely the Cartesian conception of human beings as two one-sided things, a mind and a body. Their two-way causal interaction was invoked to explain human experience and behaviour.

The greatest figures of the first two generations of twentieth-century neuroscientists, e.g. Sherrington, Eccles and Penfield, were avowed Cartesian dualists. The third generation retained the basic Cartesian structure, but transformed it into brain-body dualism: substance-dualism was abandoned, but structural dualism retained. For neuroscientists now ascribe much the same array of mental predicates to the brain as Descartes ascribed to the mind, and conceive of the relationship between thought and action, and experience and its objects, in much the same way as Descartes – essentially merely replacing the mind by the brain. The central theme of our book was to demonstrate the incoherence of brain/body dualism, and to disclose its misguided crypto-Cartesian character. Our constructive aim was to show that an Aristotelian account, with due emphasis on first- and second-order active and passive abilities and their modes of behavioural manifestation, is necessary to do justice to the structure of our conceptual scheme and to provide coherent descriptions of the great discoveries of post-Sherringtonian cognitive neuroscience.[iii]

Aristotle's principle

In the book we identified a pervasive error that we called 'the mereological fallacy in neuroscience'.[iv]

Correcting this error is a *leitmotiv* (but *only* a *leitmotiv*) of our book. We called the mistake 'mereological', because it involves ascribing to parts attributes that can intelligibly be ascribed only to the wholes of which they are parts. A form of this error was pointed out around 350 BC by Aristotle,

who remarked that 'to say that the soul [the *psuche*] is angry is as if one were to say that the soul weaves or builds. For it is surely better not to say that the soul pities, learns or thinks, but that a man does these with his soul' (DA 408[b]12-15) – doing something with one's soul being like doing something with one's talents. It is mistaken to ascribe to the soul of an animal attributes that are properly ascribable only to the animal as a whole. We might call this 'Aristotle's principle'.

Our primary concern was with the neuroscientific cousin of this, namely the error of ascribing to the *brain* – a part of an animal – attributes that can be ascribed literally only to the animal as a whole. We were not the first to have noted this – it was pointed out by Anthony Kenny in his brilliant paper 'The Homunculus Fallacy' of 1971.[v] This is more properly *mereological* than Aristotle's principle, since the brain is literally a part of the sentient animal, whereas, contrary to the claims of Plato and Descartes, the soul or mind is not. In Aristotelian spirit we now observe that to say that the brain is angry is as if one were to say that the brain weaves or builds. For it is surely better to say not that the brain pities, learns or thinks, but that a man does these.[vi] Accordingly, we deny that it makes sense to say that the brain is conscious, feels sensations, perceives, thinks, knows or wants anything – for these are attributes of animals, not of their brains.

We are a little surprised to find that Professor Dennett thinks that his distinction in *Content and Consciousness* of 1969 between personal and subpersonal levels of explanations is what *we* had in mind. He there wrote, correctly, that being in pain is not a property of the brain. But his reason was that pains are 'mental phenomena' that are 'non-mechanical', whereas cerebral processes are 'essentially mechanical' (ibid., p. 91). The contrast *we* drew between properties of wholes and properties of parts is not between what is non-mechanical and what is mechanical. It is the bracket clock as a whole that keeps time, not its fusée – although the process of keeping time is wholly mechanical. It is the aeroplane that flies, not its engines – although the process of flying is wholly mechanical. Moreover, verbs of sensation, such as 'hurts', 'itches', 'tickles' *do* apply to the parts of an animal, whose leg may hurt, whose head may itch and whose flanks may tickle (PFN 73). These attributes are, as Professor Dennett puts it, 'non-mechanical'; nevertheless they *are* ascribable to parts

of an animal. So the mereological point we made is quite different from Professor Dennett's distinction between personal and sub-personal levels of explanation, and, applied to animals, is quite different from his distinction between what is 'mechanical' and what is not.

Why should one accept Aristotle's principle and its neuroscientific cousin? Why should we discourage neuroscientists from ascribing *consciousness, knowledge, perception,* etc. to the brain?

Consciousness. It is animals that are conscious or unconscious, and that may become conscious of something that catches their attention. It is the student, not his brain, who awakes and becomes conscious of what the lecturer is talking about, and it is the lecturer, not his brain, who is conscious of his students' boredom as they surreptitiously yawn. The brain is not an organ of consciousness. One sees with one's eyes and hears with one's ears, but one is not conscious with one's brain any more than one walks with one's brain.

An animal may be conscious without showing it. That is the *only* sense in which one can say, with Professor Searle, that 'the very existence of consciousness has nothing to do with behaviour'. But, *the concept* of consciousness is bound up with the behavioural grounds for ascribing consciousness to the animal. An animal does not have to exhibit such behaviour in order for it *to be* conscious. But only an animal to which such behaviour *can intelligibly be ascribed* can also be said, *either truly or falsely*, to be conscious. It makes no sense to ascribe thought to a chair or an oyster, because there is no such thing as a chair or oyster behaving thoughtfully. The 'ontological question' – the question of truth – presupposes the antecedent determination of the question of sense. Agreement on the behavioural grounds for ascription of consciousness, i.e. on what *counts* as a manifestation of consciousness, is a precondition for scientific investigation into the neural conditions for being conscious. Otherwise one could not even identify what one wants to investigate.

Professor Searle insists that consciousness is a property of the brain. Sherrington, Eccles and Penfield, being Cartesians, thought it to be a property of the mind. What recent neuroscientific experiment can Professor Searle cite to show that it is *actually* a property of the brain? After all, the only thing neuroscientists *could* discover is that certain neural states are inductively well-correlated

with an *animal's* being conscious. Is Professor Searle's claim then a conceptual insight? No – for that is not the way the concept of *being conscious* is deployed. It is human beings (and other animals), not their brains (or their minds), that fall asleep and later awaken, that are knocked unconscious and later regain consciousness. So is it a linguistic recommendation: namely, that when a human being's brain is in a state that is inductively well-correlated with the human being's being conscious, we should describe his brain as being conscious too? This is a convention we could adopt. We could introduce this secondary use of 'to be conscious'. It is necessarily parasitic on the primary use that applies to the human being as a whole. It is, however, difficult to see anything that recommends it. It is certainly not needed for the sake of clarity of neuroscientific description, and it adds nothing but an empty form to existing neuroscientific explanation.

Knowledge. Knowledge comprises abilities of complex kinds. The identity of an ability is determined by what it is an ability to do. The simplest grounds for ascribing an ability to an animal is that it engages in corporeal activities that manifest its abilities. The more complex the ability, the more diverse and diffuse the grounds. If an animal knows something, it can act and respond to its environment in ways that it cannot if it is ignorant; if it does so, it manifests its knowledge. The brain can be said to be the *vehicle* of these abilities, but what this means is that in the absence of the appropriate neural structures the animal would not be able to do what it can do. The neural structures in the brain are distinct from the abilities *the animal* has, and the operations of these structures are distinct from the exercise of the abilities *by the animal*. In short, the knower is also the doer, and his knowing is exhibited in what he does.

We pointed out that J.Z. Young, like so many neuroscientists, held that the brain contains knowledge and information 'just as knowledge and information can be recorded in books or computers'. Contrary to what Professor Dennett asserts, there is no question but that Young meant these terms to be understood in their customary sense – for it is in their customary sense that knowledge and information can be recorded in books. Professor Dennett says that we did nothing to establish that there is no concept of knowledge or information such that it cannot be said to be

encoded in both books and brains (D 13). In fact we did discuss this (PFN 152f.). But we shall explain again.

A code is a system of encrypting conventions parasitic on language. A code is not a language. It has neither a grammar nor a lexicon (cf. Morse code). Knowledge is not *encoded* in books, unless they are written in code. One can encode a message only if there is a code in which to do so. There is a code only if encoders and intended decoders agree on encoding conventions. In this sense there isn't, and couldn't be, a neural code. In the sense in which a book contains information, the brain contains none. In the sense in which a human being possesses information, the brain possesses none. That information can be derived from features of the brain (as dendrochronological information can be derived from a tree trunk) does not show that information is encoded in the brain (any more than it is in the tree trunk).

So, in the ordinary sense of 'knowledge', there can be no knowledge recorded, contained in, or possessed by the brain. Professor Dennett then switched tack, and recommended that we attend to the cognitive scientific literature on *extensions* of the term 'knowledge' that might allow knowledge, in an extended sense, to be ascribed to the brain. And he recommended to our attention Chomsky's attempt to explain an extended concept of knowledge, namely 'cognizing', according to which human beings, and even neonates, cognize the principles of universal grammar.[vii] According to Chomsky, someone who cognizes cannot tell one what he cognizes, cannot display the object of his cognizing, does not recognize what he cognizes when told, never forgets what he cognizes (but never remembers it either), has never learnt it and could not teach it. Apart from that, cognizing is just like knowing! Does *this* commend itself as a model for an intelligible extension of a term?

Perception: The perceptual faculties are powers to acquire knowledge by the use of one's sense organs. An animal uses its eyes in glancing at, watching, peering at, and looking at things. It is thus able to discriminate things that are coloured, that have distinctive shapes and movements. It exhibits its visual acumen in what it does in response to what it sees. It would not have these perceptual powers or be able to exercise them but for the proper functioning of appropriate parts of its

brain. However, it is not the cerebral cortex that sees, but the animal. It is not the brain that moves closer to see better, looks through the bushes and under the hedges. It is not the brain that leaps to avoid a predator seen, or charges the prey it sees – it is the perceiving animal. In short, the perceiver is also the actor.

In *Consciousness Explained*, Professor Dennett ascribed psychological attributes to the brain (e.g. he asserted that it is conscious, gathers information, makes simplifying assumptions, makes use of supporting information, and arrives at conclusions (CE 142-4)). He now avers that ascribing such predicates to the brain is indeed mistaken. Nevertheless, he holds, it is theoretically fruitful, and consistent with accepting the erroneous character of attributing predicates of wholes to their parts, to extend the psychological vocabulary, duly attenuated, from human beings and other animals to (a) computers and (b) parts of the brain. Indeed, he apparently holds that there is no difference of moment between these two extensions. But there is a difference. Attributing psychological properties to computers is mistaken, but does not involve a mereological fallacy. Attributing psychological properties to the brain or its parts is mistaken and does involve a mereological fallacy. Taking the brain to be a computer and ascribing psychological properties to it or its parts is doubly mistaken. Let me explain.

It is true that we do, in casual parlance, say that computers remember, that they search their memory, that they calculate, and sometimes, when they take a long time, we jocularly say that they are thinking things over. But this is merely a *façon de parler*. It is not a literal application of the terms 'remember', 'calculate' and 'think'. Computers are devices designed to fulfil certain functions for us. We can store information in a computer, as we can in a filing cabinet. But filing cabinets cannot remember anything, and neither can computers. We use computers to produce the results of a calculation – just as we used to use a slide-rule or cylindrical mechanical calculator. Those results are produced without anyone or anything literally calculating – as is evident in the case of a slide-rule or mechanical calculator. In order literally to calculate, one must have a grasp of a wide range of concepts, follow a multitude of rules that one must know, and understand a variety of operations.

Computers do not and cannot.

Professor Dennett suggests that 'it is an empirical fact ... that *parts* of our brains engage in processes that are *strikingly like* guessing, deciding, believing, jumping to conclusions, etc. and it is *enough* like these personal level behaviors to warrant stretching ordinary usage to cover it' (D 10). He agrees that it would be mistaken to 'attribute *fully-fledged* belief', decision, desire or pain to the brain. Rather, 'just as a young child can *sort of* believe that her Daddy is a doctor ... , so ... some part of a person's brain can *sort of* believe that there is an open door a few feet ahead' (D 11).

This is part of what Professor Dennett characterizes as 'the intentional stance' – a research methodology that supposedly helps neuroscientists to explain the neural foundations of human powers. He claims that adoption of the intentional stance has accomplished 'excellent scientific work ... generating hypotheses to test, articulating theories, analysing distressingly complex phenomena into their more comprehensible parts' (D 10). It seems committed to the idea that some parts of the brain 'sort of believe', that others *sort of decide*, and yet others *sort of oversee* these activities. All this, presumably, is supposed to *sort of explain* what neuroscientists want to explain. But if the explananda are uniformly sorts of believings, pseudo-expectings, proto-wantings and demi-decidings, they at best only *sort of make sense*, and presumably are only *sort of true*. And how one can make valid inferences from such premisses is more than just sort of obscure. How precisely such premisses are supposed to *explain* the phenomena is equally obscure. For the logic of such putative explanations is altogether unclear. Does sort of believing, pseudo-believing, proto-believing or demi-believing something furnish a part of the brain with a reason for acting? Or only a sort of reason? – for a sort of action? When asked whether parts of the brain are, as Dennett puts it, 'real intentional systems', his reply is 'Don't ask' (D 12).[viii]

Cognitive neuroscientists ask *real* questions – they ask *how* the prefrontal cortices are involved in human thinking, *why* re-entrant pathways exist, *what* precisely are the roles of the hippocampus and neocortex in a human being's remembering. Being told that the hippocampus sort of remembers for a short while and that the neocortex has a better sort of long-term memory provides

no explanation whatsoever. No well-confirmed empirical theory in neuroscience has emerged from Dennett's explanations, for ascribing 'sort of psychological properties' to parts of the brain does not *explain* anything. We shall revert to this when we discuss Sperry and Gazzaniga's account of commissurotomy. Not only does it not explain, it generates further incoherence.[ix]

We agree with Professor Dennett that many of a child's beliefs are beliefs in an attenuated sense. A little girl's grasp of the concept of a doctor may be defective, but she will rightly say 'Daddy is a doctor', and reply to the question 'Where is the doctor?' by saying 'In there ☞ (pointing to Daddy's office)'. So she can be said to believe, in an attenuated sense, that her father is a doctor. She satisfies, in her verbal and deictic behaviour, *some* of the normal criteria for believing that her father is a doctor (but also satisfies some of the criteria for lacking this belief). But there is no such thing as a part of a brain asserting things, as the child does, answering questions, as the child does, or pointing at things, as the child does. So in the sense in which, in her verbal and deictic behaviour, the child can manifest *rudimentary* belief, a part of a brain can no more do so than the whole brain can manifest fully-fledged belief. Or can Professor Dennett suggest an *experimentum crucis* that will demonstrate that her prefrontal cortex sort of believes that the cat is under the sofa?

The child can also exhibit rudimentary belief in her non-verbal behaviour. If she sees the cat run under the sofa and toddles over to look for it, then she can be said to think the cat is under the sofa. But brains and their parts cannot *behave*, cannot toddle over to the sofa, cannot look under it, and cannot look nonplussed when there is no cat there. Brain parts can neither voluntarily act nor take action. Unlike the child, brain parts cannot satisfy *any* of the criteria for believing something, even in a rudimentary sense. Brains (and their parts) can only 'sort of believe' in the sense in which they are 'sort of oceans' (since there are brain-waves), and are 'sort of weather-systems' (since there are brainstorms). The similarity between a brain and an ocean is at least as great as the similarity of brain processes to human beings' believings, decidings, or guessings. (After all, both brains and oceans are grey, have wrinkles on their surface, and have currents running through them.)

Is the mereological fallacy mereological ?

Professor Searle objects that what we characterize as a paradigm of a mereological fallacy, i.e. the ascription of psychological attributes to the brain, is no such thing, for the brain is not a part of a person, but rather a part of a person's body. This, we think, is a red-herring. The dictum of Wittgenstein that we quoted was 'Only of a *human being* and of what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf', etc. The brain *is* a part of the human being.

Professor Searle suggests that if ascribing psychological attributes to the brain really were a mereological error, then it would vanish if one ascribed them to what he calls 'the rest of the system' to which the brain belongs. He thinks that the 'rest of the system' is the body that a human being *has*. He observes that we do not ascribe psychological predicates to the body one has. With the striking exception of verbs of sensation, the latter point is correct. We do not say 'My body perceives, thinks, or knows'. However, 'the system' to which the human brain can be said to belong is *the human being*. The human brain is a part of the human being, just as the canine brain is a part of a dog. My brain, the brain I have, is as much a part of me – of the living human being that I am – as my legs and arms are parts of me. But it is true that my brain can also be said to be a part of my body.

How is this to be explained? Our talk of our mind is largely *non-agential*, *idiomatic* talk of our rational powers of intellect and will, and of their exercise. Our talk of our body is talk of our corporeal properties. To speak of my body is to speak of corporeal features *of the human being that I am* – features pertaining to appearance (an attractive or ungainly body), to aspects of health and fitness (a diseased or healthy body), and, very strikingly, to sensation (my body may ache all over, just as my leg may hurt and my back may itch).[x] But knowing, perceiving, thinking, imagining, etc. are not corporeal features of human beings and are therefore not ascribable to the body a human being has, any more than they are ascribable to the brain that a human being has. Human beings are not *their bodies*. Nevertheless they *are* bodies, in the quite different sense of being a particular kind of

sentient spatio-temporal continuant – *homo sapiens*; and the brain is a part of the living human being, as are the limbs.[xi] It is not, however, a conscious, thinking, perceiving part – and nor is any other part of a human being. For these are attributes of the human being as a whole.

Nevertheless, Professor Searle has noted an interesting feature of our corporeal idiom. Human beings are persons – that is, they are intelligent, language-using animals, are self-conscious, possess knowledge of good and evil, are responsible for their deeds, and are bearers of rights and duties. To be a person is to possess such abilities as qualify one for the status of a moral agent. We would probably not say that the brain is part of the person, but rather that it is part of the person's body, whereas we would not hesitate to say that Jack's brain is a part of Jack, part of *this* ☞ human being, just as his legs and arms are parts of Jack. Why? Perhaps because 'person' is, as Locke stressed 'a forensic term', but not a substance-name. So, if we use the term 'person' in contexts such as this, we indicate thereby that we are concerned primarily with human beings qua possessors of those characteristics that render them persons, in relative disregard of corporeal characteristics. Perhaps the following analogy will help: London is a part of the UK; the UK belongs the European Union, but London does not belong to the EU. That does not prevent London from being part of the UK. So too Jack's being a person does not prevent his brain being part of him.

Qualia

In our discussion of consciousness (PFN chaps. 9-12), we argued that characterizing the domain of the mental by reference to the 'qualitative feel' of experience is misconceived (PFN chap. 10). But *pace* Professor Searle (S 4, 13), we did not deny the existence of qualia on the grounds that if they did exist, they would exist in brains. If, *per impossibile* psychological attributes were all characterized by their 'qualitative feel', they would still be attributes of human beings, not of brains.

A quale is supposed to be 'the qualitative feel of an experience' (Chalmers)[xii], or it is such a thing as 'the redness of red or the painfulness of pain' (Crick)[xiii]. Qualia are 'the simple sensory qualities to be found in the blueness of the sky or the sound of a tone' (Damasio)[xiv]; or 'ways it feels

to see, hear and smell, the way it feels to have a pain' (Block)[xv], According to Professor Searle, conscious states are 'qualitative in the sense that for any conscious state ... there is something that it qualitatively feels like to be in that state'.[xvi] According to Nagel for every conscious experience, 'there is something it is like for the organism to have it'.[xvii] These various explanations *do not amount to the same thing*, and it is questionable whether a coherent account emerges from them.

Professor Searle remarks that there is a qualitative feel to a pain, a tickle and an itch. To this we agree – in the following sense: sensations, we remarked (PFN 124), have phenomenal qualities (e.g. burning, stinging, gnawing, piercing, throbbing); they are linked with felt inclinations to behave (to scratch, assuage, giggle or laugh); they have degrees of intensity which may wax or wane.

When it comes to perceiving, however, we noted that it is problematic to characterize what is meant by 'the qualitative character of experience'. Specifying what we see or smell, or, in the case of hallucinations, what it seems to us that we see or smell, requires specification of an object. Visual or olfactory experiences and their hallucinatory counterparts are individuated by what they are experiences or hallucinations of. Seeing a lamp-post is distinct from seeing a postbox, smelling lilac is different from smelling roses, and so too are the corresponding hallucinatory experiences that are described in terms of their seeming to the subject to be like their veridical perceptual counterpart.[xviii]

To be sure, roses do not smell like lilac – what roses smell like is different from what lilac smells like. Smelling roses is quite different from smelling lilac. But the qualitative character of smelling roses does not smell of roses, any more than the qualitative character of smelling lilac smells of lilac. Smelling either may be equally pleasant – in which case the qualitative character of the smelling may be exactly the same, even though what is smelled is quite different. Professor Searle, we suggest, confuses what the smells are like with what the smelling is like.

Seeing a lamp-post does not normally feel like anything. If asked 'What did it feel like to see it?', the only kind of answer is one such as 'It didn't feel like anything in particular – neither pleasant nor unpleasant, neither exciting nor dull'. *Such* epithets – 'pleasant', 'unpleasant', 'exciting', 'dull' – *are* correctly understood as describing the 'qualitative character of the experience'. In this sense,

many perceptual experiences have no qualitative character at all. *None* are individuated by their qualitative feel – they are individuated by their object. And if we are dealing with a hallucination, then saying that the hallucinated lamp-post was black is still the description of the object of the experience – its 'intentional object' in Brentano's jargon (which Professor Searle uses). The quality of the hallucinatory experience, on the other hand, is probably: *rather scary*.

Contrary to what Professor Searle suggests, we did not argue that 'if you do not define qualia as a matter of pleasantness or unpleasantness then you will have to individuate the experience by its object' (S 14). Our argument was that we *do* individuate experiences and hallucinations by their objects – which are specified by the answer to the question 'What was your experience (or hallucination) an experience (or hallucination) *of* ?' Of course, the object need not be the cause, as is evident in the case of hallucinations. But, we insisted, the qualitative character of the experience should not be confused with the qualities of the object of the experience. That what one sees when one sees a red apple is red and round does not imply that one enjoyed a red, round visual experience. That what one seems to see when one hallucinates a red apple is red and round does not imply that one enjoyed a red, round visual hallucination. 'What did you see (or hallucinate)?' is one question, 'What was it like to see what you saw (or hallucinate what you hallucinated)?' another. One does not individuate perceptual experiences by their qualitative character. Professor Searle holds that there is something it is like to think that 2+2=4, which is presumably quite different from what it is like to think that 3+3=6. Can Professor Searle tell us precisely what it is like? Surely, the question 'What is it *like* to think that 2+2=4?'is misconceived. One does not individuate *thinking* or *believing* by any qualitative character, but by what is thought or believed to be so. These are simple truths; but they seem to have been overlooked.

Enskulled brains

Professor Searle suggests that human beings are 'embodied brains' (S 16f.). According to his view, the reason why we can say both 'I weigh 160 lbs' and 'My body weighs 160 lbs' is that what makes it

the case that I weigh 160 lbs is that my body does. But I, it seems, am strictly speaking no more than an embodied (enskulled) brain. I *have* a body, and I am *in* the skull *of* my body. This is a materialist version of Cartesianism. One major reason why we wrote our book was the firm belief that contemporary neuroscientists, and many philosophers too, still stand in the long, dark shadow of Descartes. For while rejecting the immaterial substance of the Cartesian mind, they transfer the attributes of the Cartesian mind to the human brain instead, leaving intact the whole misconceived structure of the Cartesian conception of the relationship between mind and body. What we were advocating was that neuroscientists, and even philosophers, leave the Cartesian shadow lands and seek out the Aristotelian sunlight, where one can see so much better.

If I were, *per impossibile*, an embodied brain, then I would have a body – just as the Cartesian embodied mind has a body. But I would not *have* a brain, since brains do not have brains. And in truth my body would not weigh 160 lbs, but 160 lbs less 3 lbs – which is, strictly speaking, what *I* would weigh. And I would not be 6 foot tall, but only 7 inches tall. Doubtless Professor Searle will assure me that I am my-embodied-brain – my brain *together* with my body. But that does not get us back on track. For my brain together with my brainless body, taken one way, is just my cadaver; taken another way, it is simply *my body*. But I am not my body, not the body *I have*. Of course, I am *a* body – the living human being that stands before you, a particular kind of sentient spatio-temporal continuant that possesses intellect and will and is therefore a person. But I am no more *my body* than I am *my mind*. And I am not an embodied brain either. It is mistaken to suppose that human beings are 'embodied' at all – that conception belongs to the Platonic, Augustinian and Cartesian tradition that should be repudiated. It would be far better to say, with Aristotle, that human beings are *ensouled* creatures ('*empsuchos*') – animals endowed with such capacities that confer upon them, in the form of life that is natural to them, the status of persons.

Neuroscientific research

Professor Searle claims that central questions in neurobiological research would be rejected as

meaningless if our account of the conceptual structures deployed were correct. So, he suggests, 'the central question in vision, how do neurobiological processes ... cause conscious visual experiences, could not be investigated by anyone who accepted [our] conception' (S 20). Our conception, he avers, 'can have potentially disastrous scientific consequences' (ibid.).

Research on the neurobiology of vision is research into the neural structures that are causally necessary for an animal to be able to see, and into the specific processes involved in its seeing. That we deny that visual experiences occur in the brain, or that they are characterized by qualia, affects this neuroscientific research programme only in so far as it averts futile questions that could have no answer. We gave numerous examples, e.g. the binding problem (Crick, Kandel and Wurtz), or the explanation of recognition by reference to the matching of templates with images (Marr), or the suggestion that perceptions are hypotheses of the brain that are conclusions of unconscious inferences it makes (Helmholtz, Gregory, and Blakemore). Our contention that it is the animal that sees or has visual experiences, not the brain, and Professor Searle's contention that it is the brain, not the animal, are conceptual claims, not empirical ones. The issue is none the less important for all that, but it should be evident that what we said does not *hinder* empirical investigation into the neural processes that underpin vision. Rather, it guides the description of the results of such investigations down the highroads of sense.

Professor Dennett holds that our refusal to ascribe psychological attributes (even in an attenuated sense) to anything less than an animal as a whole is retrograde and unscientific. In his view, 'the poetic license granted by the intentional stance eases the task' of explaining how the functioning of parts contributes to the behaviour of the animal (D 12).

We note first that *poetic license* is something granted to poets for purposes of poetry, not for purposes of empirical precision and explanatory power. Secondly, ascribing cognitive powers to parts of the brain provides only the semblance of an explanation where an explanation is still wanting. So it actually blocks scientific progress. Sperry and Gazzaniga claim that in cases of commissurotomy, the bizarre behaviour of subjects under experimental conditions of exposure to pictured objects is

explained by the fact that one hemisphere of the brain is ignorant of what the other half can see. The hemispheres of the brain allegedly know things and can explain things, and, because of the severance of the corpus callosum, the right hemisphere allegedly cannot communicate to the left hemisphere what it sees. So the left hemisphere must generate its own interpretation of why the left hand is doing what it is doing.[xix] Far from explaining the phenomena, this masks the absence of any substantial explanation by redescribing them in misleading terms. The dissociation of functions normally associated is indeed partially explained by the severing of the corpus callosum and by the localization of function in the two hemispheres. *That* is now well known, but currently available explanation goes no further. It is an illusion to suppose that anything whatsoever is added by ascribing knowledge, perception, and linguistic understanding (sort of, or otherwise) to the hemispheres of the brain.

Professor Dennett holds that in refusing to acknowledge that there are maps and images in the brain, we are obfuscating empirical research (D 14f.). For, in his view, the determinate patterns of neural stimulation in the 'visual' striate cortex have the geometric properties of images, and whether they function as images is an empirical question. Indeed, he asserts that these patterns of neural firings onto which elements of the visual field can be mapped *are maps*, and that the brain *makes use of them as maps* (D 15, fn 19). This, we suggest, is mistaken. The discoveries of Hubel and Wiesel are admirable. But to discover a mapping is not to discover a map. To *use* a map as a map, there has to *be* a map – and there are none in the brain; one has to be able to read the map – but brains lack eyes and cannot read; one has to be familiar with the projective conventions of the map (e.g. cylindrical, conic, azimuthal) – but there are no projective *conventions* regarding the mappings of features of the visual field onto the neural firings in the 'visual' striate cortex; and one has to use the map to guide one's behaviour – one's perambulations or navigations – which are not activities brains engage in. Far from obfuscating the results of empirical research, these conceptual truths clarify them and restrain judicious neuroscientists within the bounds of sense. They do not derogate from their impressive discoveries, or stand in the way of explanatory, testable, hypotheses.

In general, the conceptual criticisms in our book do no more than peel away layers of

conceptual confusion from neuroscientific research and clarify the conceptual forms it presupposes. This cannot impede the progress of neuroscience.

Appendices (not delivered)

1. *Grounds, truth-conditions and the conditions for a practice*

The grounds on which one may ascribe psychological attributes to a creature may be logical (constitutive) or inductive. Inductive grounds presuppose non-inductive identifications, otherwise there can be no inductive correlations. Non-inductive identifications, in such cases, have as their prerequisite logically good evidence – criteria – for the ascription of the psychological predicate. That such-and-such behaviour is a criterion for the application of such a predicate is partly constitutive of its meaning. We learn the use of these predicates by learning what forms of behaviour in what kinds of circumstances warrant their ascription to others. We learn to use them in our own case, sometimes as a partial substitute for our natural behaviour ('It hurts', 'I want'), sometimes as grafted onto already acquired linguistic behaviour ('I think that'). It is no inductive coincidence that pain is manifest in shrieks of agony, that perception is exhibited in discriminatory responses to perceptibilia, or that affection is shown in affectionate behaviour.

But, of course, one may be in pain and not manifest it, perceive and not show it, feel affection and not demonstrate it. We agree with Professor Searle that it would be mistaken to conflate the grounds for ascribing a psychological predicate with what they are grounds for, and to fail to distinguish the evidence for being in pain from being in pain. What puzzles us is why he thought we disagree. If a truth-condition of being in pain is: *being in pain*, then to be sure, pain behaviour is altogether distinct from the truth-condition of being in pain. So too, it is true that A is in pain if and only if it is a fact that he is in pain, i.e. if he is, in fact, in pain, or – more plainly – if he is in pain. We would be the last to deny this.

We are defending the claim that it only makes sense to ascribe a psychological attribute G to a being if it makes sense to describe it as behaving in a manner that exhibits G. That does not imply that it is only G if it is manifesting G. We are not defending any form of behaviourism, and certainly not claiming that behaving in a manner that is a criterion for being G is what being G consists in (PFN 82, fn. 35).

It has been suggested that we conflate the conditions under which a linguistic practice can be engaged in – or, as Professor Searle puts it, *a language game can be played* – with the existence of the phenomena described by the expressions that are part of the practice. We agree that it is important to distinguish these. The rules of tennis are one thing, whether anyone is playing tennis is another, and the conditions under which it can be played to any purpose a third. (No one would try to play tennis on the moon, although the rules make no mention of gravitational fields.) Our only concern in our book is with the norms of correct use of the psychological vocabulary. What the conceptual commitments of our psychological vocabulary are is one question, whether a creature is suffering, perceiving, thinking is another, and what the background conditions for playing the language-game with psychological predicates are is a third. Our concern was solely with the first question. We agree with Professor Searle that it would be erroneous to say that the conditions for the successful operation of the language-game are conditions for the existence of the phenomena described. That certain behaviour in circumstances is a criterion for the ascription of a predicate G to an animal is not a condition for it to be G. Nor is it a condition for the existence of the practice of using 'G'. It is one of the rules constitutive of the practice. That an animal display G-behaviour is not a condition for the existence of the practice of using 'G' – even if no one is G and no one is displaying G-behaviour, one could still assert 'I am not G', or ask 'Are you G?' and get a negative answer. What *is* a condition of the existence of the practice is that behaviour in appropriate circumstances display such regularities as warrant applying or withholding the predicate 'G'. If behaviour became confused, if tears and laughter became systematically intermingled, if smiles and scowls alternated rapidly for no apparent reason, if declarations of intent were never followed by action, and so forth – then various linguistic practices with certain psychological predicates would disintegrate, and the language-game, as Professor Searle says, would not be played. With this we agree. But confusing the rules of the game, the existence of the phenomena described, and the conditions in which the game can be played is not an error we committed.

2. *Location*

The question of whether the brain is a possible subject of psychological attributes is distinct from the question of whether the brain is the locus of those psychological attributes to which a corporeal location can intelligibly be assigned. We agree with Professor Searle on the importance of distinguishing these questions, and in fact we explicitly distinguished them (PFN 122f., 179).

We also agree that our reasons for denying that the brain can be the subject of psychological attributes do not show that the brain is not the locus of such attributes to which it makes sense to assign a corporeal location. Nor were they meant to. Our view is that sensations such as pains and itches can be assigned a location, whereas thinking, believing, deciding, and wanting, for example, cannot. The answer to the questions 'Where did you think of that?', 'Where did he acquire that strange belief?', 'Where did she decide to get married?' is *never* 'In the prefrontal lobes, of course'. The location of a human being's thinking, recollecting, seeing, deciding is where the human being is when he thinks, etc. Which part of his brain is involved in his doing so is a further, important question about which neuroscientists are gradually learning more. But they are not learning where thinking, recollecting, or deciding occur – they are discovering which parts of the cortex are causally implicated in a human person's thinking, recollecting, deciding.

Of course, thinking about something, deciding to do something, seeing something, are, as Professor Searle rightly says, real events – they really happen somewhere, somewhen, in the world. I thought up that argument in the library, and decided how to phrase it in my study; I saw Jack when I was in the street and I listened to Jill's recital in the concert hall. Professor Searle suggests that the question 'Where do mental events occur?' is no more philosophically puzzling than the question: 'Where do digestive processes occur?' So, he argues, digestive processes occur in the stomach, and consciousness occurs in the brain. This is mistaken. Being conscious, as opposed to unconscious, being conscious of something, as opposed to not noticing it or not attending to it, do not occur *in* the brain at all. Of course, they occur *because of certain events in the brain*, without which a human being would not have regained consciousness or had his attention caught. 'Where did you become conscious of the sound of the clock?' is to be answered by specifying where I was when it caught my attention, just as 'Where did you regain consciousness?' is to be answered by specifying where I was when I came round.

Both digesting and thinking are predicated of animals. But it does not follow that there are no logical differences between them. The stomach can be said to be digesting food, but the brain cannot be said to be thinking. The stomach is the digestive organ, the brain is not an organ of thought. If one opens the stomach, one can see the food being digested there. But if one wants to see thinking going on, one should look at the *Penseur* (or the surgeon operating, or the chess player playing or the debater debating) not at his brain. All his brain can show is what goes on there while he is thinking. We ascribe length, strength and having cracks to steel girders. But it does not follow that *length* and *strength* have the same logical character; and one can ask where the crack is, but not where the strength is.

So, sensations, such as pains, *are* located in our bodies. But Professor Searle holds that they are all *in the brain*. It is, as he admits, counterintuitive – after all we complain of stomach-ache, of gout in our foot or arthritis in our knees. Nevertheless, he claims, the brain creates a body image, and the pain that we describe as being in the foot, and which we assuage by rubbing the foot, is an awareness-of-the-pain-as-in-my-foot, which is in the body-image that is in one's brain. It is interesting that Descartes took a very similar view, remarking that 'pain in the hand is felt by the soul not because it is present in the hand but because it is present in the brain'.[xx] The advantage of his account, Professor Searle suggests, is that it means that we can describe the phenomenon of phantom pain without the absurdity of suggesting that the pain is in physical space, in the bed or underneath the sheet. But that absurdity, he holds, is what we are committed to by claiming that pains are in the body. We agree on the absurdity, but deny that we are committed to it.

There are many locative uses of 'in', some spatial, some non-spatial ('in the story', 'in October', 'in committee'). Among spatial uses, there are many different kinds, depending on what is in what (PFN 123f.). We agree with Professor Searle that if there is a coin in my jacket pocket, and if

my jacket is in the dresser, then there is coin in the dresser. But not all spatial locative uses of 'in' are thus transitive. If there is a hole in my jacket and the jacket is in the wardrobe, it does not follow that there is a hole in the wardrobe. In the case of the jacket and the coin, we are concerned with spatial relations between two independent objects, but not in the case of the jacket and the hole. Similarly, if there is a crease in my shirt, and my shirt is in the suitcase, it does not follow that there is a crease in the suitcase. The coin may be taken out of the jacket pocket, and the shirt may be taken out of the suitcase, but the hole cannot be taken out of the pocket – it has to be sewn up, as the crease has to be ironed out, not taken out.

The use of 'in' with respect to the location of sensations is not like the coin, but more like the hole (though still different). A pain is not a substance. If I have a pain in my foot, I do not stand in any *relation* to a pain – rather, my foot hurts *there* ☞, and I can point to the place that hurts, which we call 'the location of the pain'. In the case of the phantom limb, it feels to the sufferer just as if he still has the limb that has been amputated, and he avows a pain in the illusory limb. It seems to him just as if his leg were hurting, although he has no leg. We agree with Professor Searle that it is not the bed that hurts nor is the pain the amputee feels under the sheet. That he feels the pain where his leg would have been, and that his leg would have been under the sheet, do not imply that there is a pain under the sheet, any more than his having a pain in his unamputated leg and his leg being in his boot implies that he has a pain in his boot. Indeed, we agree with Professor Searle about the phenomena, and disagree only over its description. We do not think that there are body-images in the brain, and wonder what evidence there is for their existence – after all, one cannot find body images if one opens up the brain of a human being. What Professor Searle refers to is that physiological methods, beginning with those of Sherrington, have been used to establish that neurons in the somato-sensory cortex can be excited in a topographical one-to-one relation with points stimulated on the surface of the body and with the spatial layout of the muscles of the limbs and trunk. But we do not understand what Professor Searle means by 'having a pain in a phenomenological phantom foot in a body image in the brain'. One *can* have pains in one's *head* – they are commonly known as headaches. But one cannot have a back-ache or a stomach ache in one's brain; or any other pain. And that is no coincidence, since there are no fibre-endings there save in the dura.

Finally, Professor Searle claims that when philosophers say that two people both have the same pain, what they mean is that they have the same type-pain, but different token-pains. This is mistaken. Peirce's type/token distinction was applied to inscriptions, and is dependent on orthographic conventions. It no more applies to pains than it does to colours. If two armchairs are both maroon, then there are two chairs of the very same colour, and not two token-colours of the same type. For how is one to individuate the different tokens? All one can say is that the first alleged token belongs to the first chair and the second to the second chair. But this is to individuate a property by reference to the pseudo-property of belonging to the substance that has it – as if properties were substances that are distinguished by means of Leibniz's law, and as if being the property of a given substance were a property that distinguishes, for example, the colour of *this* chair from the colour of *that* one. And that is absurd. The two chairs are both of the very same colour. Similarly, if two people have a splitting headache in their left temples, then they both have the very same pain. A's pain is not distinguishable from B's pain by virtue of the fact that it belongs to A, any more than the maroon colour of the first chair is distinguished from the maroon colour of the second chair by virtue of the fact that it belongs to the first. The distinction between qualitative and numerical identity does not apply to colours or to pains, and neither does the Peircean distinction between types and tokens.

Endnotes

i. It might seem that in explaining existing concepts of cognition, cogitation, perception, affection, etc. we are committed to a form of linguistic or conceptual conservatism. Nothing could be further from the truth. We are trying to understand the conceptual scheme that neuroscientists employ at the moment, not a different one they may devise in the future. It is patent from their writings that their deployment of the psychological vocabulary to characterize their explananda is intended to conform to current usage. Our concern was with the ensuing misuses and conceptual confusions. This places no constraint on conceptual innovation, which, if and when it is introduced, will be subject to the same critical scrutiny to weed out conceptual confusion, if and when it arises. For detailed discussion of this methodological issue, see our *Philosophical Foundations of Neuroscience* (Blackwell, Oxford, 2003; subsequent references will be flagged 'PFN'), pp. 379ff.

ii. To be sure, a Quinean naturalist (which Professor Dennett claims to be) holds that there is no distinction between conceptual truths and empirical ones. Rather, he would claim, the sentences of a theory face experience as a totality, and are confirmed holistically. But it is mistaken to suppose that theorems of the differential calculus were confirmed by the predictive success of Newtonian mechanics – they were confirmed by mathematical proofs. It is equally mistaken to suppose that 'vixens are female' is confirmed by the success of zoological theory or that 'bachelors are unmarried' is confirmed by the sociology of marital habits. So too, that red is darker than pink is not verified by confirmation of the theory of colour, but rather presupposed by it. (For canonical criticism of Quine on analyticity, see P.F. Strawson and H.P. Grice, 'In Defense of a Dogma', *Philosophical Review* 1956. For more recent, meticulous criticism of Quine's general position, see H.-J. Glock, *Quine and Davidson on Language, Thought and Reality* (Cambridge University Press, Cambridge, 2003). For a contrast between Quine and Wittgenstein, see P.M.S. Hacker, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy* (Blackwell, Oxford, 1996), chap.7.)

iii. The Aristotelian, anti-Cartesian, points that we emphasize are (i) Aristotle's principle, which we discuss below, (ii) Aristotle's identification of the *psuch_* with a range of capacities, (iii) that capacities are identified by what they are capacities to do, (iv) that whether a creature possesses a capacity is to be seen from its activities, (v) Aristotle's realization that whether the *psuch_* and the body are one thing or two is an incoherent question.

iv. It is, of course, not strictly a fallacy, but it leads to fallacies – invalid inferences and mistaken arguments.

v. Anthony Kenny 'The Homunculus Fallacy' in M. Grene ed. *Interpretations of Life and Mind* (Routledge, London, 1971). We preferred the less picturesque but descriptively more accurate name 'mereological fallacy' (and, correlatively, 'the mereological principle'). We found that neuroscientists were prone to dismiss as childish the fallacy of supposing that there is a homunculus in the brain, and to proceed in the next breath to ascribe psychological attributes to the brain.

vi. Not, of course, *with* his brain, in the sense in which one does things *with* one's hands or eyes, nor in the sense in which one does things with one's talents. Of course, he would not be able to do any of these things but for the normal functioning of his brain.

vii. N. Chomsky, *Rules and Representations* (Blackwell, Oxford, 1980). I discussed the matter in *Language, Sense and Nonsense* twenty years ago.

viii. Dennett here quotes from his autobiographical entry in S. Guttenplan ed. *A Companion to the Philosophy of Mind* (Blackwell, Oxford, 1994), p. 240.

ix. Of course, we are not denying that analogical extension of concepts and conceptual structures is often fruitful in science. The hydrodynamical model generated a fruitful, testable, and mathematicized, theory of electricity. Nothing comparable to this is evident in the poetic license of Dennett's intentional stance. It is evident that poetic license allows Professor Dennett to describe thermostats as *sort of believing* that it is getting too hot, and so switching off the central heating. But this adds nothing to engineering science or to the explanation of homeostatic mechanisms.

Professor Dennett asserts (D 11) that we did not address his attempts to use what he calls 'the

intentional stance' in explaining cortical processes. In fact we discussed his idea of the intentional stance at some length in PFN pp.427-31, giving seven reasons for doubting its intelligibility. Since Professor Dennett has not replied to these objections, we have, for the moment, nothing further to add on the matter.

x. It is important to note that the Cartesian conception of the body is quite mistaken. Descartes conceived of his body as an insensate machine – a material substance without sensation. But our actual conception of our body allocates sensation to the body we have – it is our body that aches all over or that itches intolerably.

xi. The human brain is part of the human being. It can also be said to be part of the body a human being is said to have. It is, however, striking that one would, we suspect, hesitate to say of a living person, as opposed to a corpse, that their body *has* two legs or, of an amputee, that their body *has* only one leg. The misleading possessive is applied to the human being and to a human corpse, but not, or only hesitantly, to the body the living human being is said to have. Although the brain is a part of the human body, we surely would not say 'my body *has* a brain' or 'My body's brain has meningitis'.

xii. D. Chalmers, *The Conscious Mind* (Oxford University Press, Oxford, 1996), p. 4

xiii. F. Crick, *The Astonishing Hypothesis* (Touchstone, London, 1995), pp.9f.

xiv. A. Damasio, *The Feeling of What Happens* (Heineman, London, 1999), p. 9

xv. Ned Block, 'Qualia' in S. Guttenplan ed. *Blackwell Companion to the Philosophy of Mind* (Blackwell, Oxford, 1994), p. 514.

xvi. Searle, *Mystery of Consciousness* (Granta Books, London, 1997), p. xiv.

xvii. T. Nagel, 'What it is like to be a bat?' repr. in his *Mortal Questions* (Cambridge University Press, Cambridge, 1979), p. 170.

xviii. Professor Searle (like Grice and Strawson) supposes that perceptual experiences are to be characterized in terms of their highest common factor with illusory and hallucinatory experiences. So all perceptual experience is, as it were, hallucination, but veridical perception is a hallucination with a special kind of cause. This, we think, is mistaken.

xix. G. Wolford, M.B. Miller, and M. Gazzaniga, 'The left hemisphere's role in hypothesis formation', *Journal of Neuroscience*, 20 (2000), RC 64 (1-4), p. 2.

xx. Descartes, *Principles of Philosophy* I - 46, 67, and especially IV - 196.