

Nicholas Humphrey James Arthur Memorial Lecture, American Museum of Natural History, New York, 1987. (Reprinted in "The Mind Made Flesh", pp. 65-85, OUP, 2002)

THE USES OF CONSCIOUSNESS¹

My talk tonight will have a hero. I give you Denis Diderot – the 18th century French philosopher novelist, aesthete, social historian, political theorist and editor of the *Encyclopaedia*. It's hard to see how he had time, but alongside everything else Diderot wrote a treatise called the *Elements of Physiology* – a patchwork of thoughts about animal and human nature, embryology, psychology and evolution. And tucked into this surprising work is this remark: "If the union of a soul to a machine is impossible, let someone prove it to me. If it is possible, let someone tell me what would be the effects of this union."²

Now, replace the word "soul" with "consciousness", and Diderot's two thought-questions become what are still the central issues in the science of mind. Could a *machine* be conscious? If it were conscious, what *difference* would it make?

The context for those questions is not hard to guess. Diderot was appalled by and simultaneously fascinated by the dualistic philosophy of René Descartes. "A tolerably clever man," Diderot wrote, "began his book with these words: '*Man, like all animals, is composed of two distinct substances, the soul and the body.*' . . . I nearly shut the book. O! ridiculous writer, if I once admit these two distinct substances, you have nothing more to teach me. For you do not know what it is that you call soul, less still how they are united, nor how they act reciprocally on one another."³

Ridiculous it may have been. But fifty years later, the young Charles Darwin was still caught up with the idea: "The soul," he wrote in one of his early notebooks, "by the consent of all is super-added."⁴

This is one issue that the philosophy of mind has now done something to resolve. First has come the realisation that there is no need to believe that consciousness is in fact something distinct from the activity of the physical brain. Rather, consciousness should be regarded as a "surface feature" of the brain, an emergent property that arises out of the combined action of its parts. Second – and in some ways equally important – has come the realisation that the human brain itself *is* a machine. So the question now is not *could* a machine be conscious or have a soul: clearly it could – I am such a machine, and so are you. Rather the question is what *kind* of machine could be conscious. How much more and how much less would a conscious machine have to resemble the human brain – nerve cells, chemicals and all? The dispute has become one between those who argue that it's simply a

matter of having the appropriate "computer programs", and those who say it's a matter of the "hardware" too.

So called "functionalists", such as Daniel Dennett, argue that if a machine has whatever it takes at the level of the functional architecture for it to behave in all those ways that human beings do, then any such machine must by definition be conscious. But the "non-functionalists" (or as they are sometimes called "mysterians"), such as David Chalmers, argue that the mysterious *quality* of consciousness would still be missing.

This is an interesting dispute (see chapter x). And yet I'd say it clearly jumps the gun. It is all very well to discuss whether a machine which fulfills in every respect our *expectations* of how a conscious being *ought to behave* would actually be conscious. But the major question is still unresolved: what exactly *are* our expectations, and how might we account for them? In short, what do we think consciousness *produces*? If a machine could be united to a soul, what effects – if any – would it have?

When Diderot asked this crucial question, I think it is obvious he was asking rhetorically for the answer "None". A machine, he was prepared to imagine, might have a soul – and yet for all practical purposes it would be indistinguishable from a machine without one:

“What difference,” he went on to ask, “between a sensitive and living pocket watch and a watch of gold, of iron, of silver and of copper? If a soul were joined to the latter, what would it produce therein?”⁵ Presumably, as a time-keeper – and that, after all, is what a watch does best – the watch would be just the same watch it was before: the soul would be no *use* to it, it wouldn't *show*.

I would not necessarily want to pin on to Diderot the authorship of the idea of the functional impotence of souls. But whenever it came, and whether or not Diderot got there, the realisation that human consciousness itself might actually be useless was something of a breakthrough. I remember my own surprise and pleasure with this "naughty" idea, when I first came across it in the writings of the Behaviourist psychologists. There was J.B. Watson, in 1928, arguing that the science of psychology need make no reference to consciousness: “The behaviourist sweeps aside all mediaeval conceptions. He drops from his scientific vocabulary all subjective terms such as sensation, perception, image, desire, and even thinking and emotion.”⁶

And there, as philosophical back-up, was Wittgenstein, arguing that concepts referring to internal states of mind have no place in the "language game".⁷ If nothing else, it was an idea to tease one's school-friends with. "How do I know that what I experience as the colour red, isn't what you experience as green? How do I know that you experience anything at all? You might be an unconscious zombie."

A naughty idea is, however, all that it amounts to: an idea which has had a good run, and now can surely be dismissed. I shall give two reasons for dismissing it. One is a kind of Panglossian argument, to the effect that whatever exists as a consequence of evolution must have a function. The other is simply an appeal to common-sense. But before I give either, let me say what I am *not* dismissing: I am not dismissing the idea that consciousness is a second-order and in some ways inessential process. In certain respects the behaviourists may have been right.

Diderot gives a nice example of *unconscious* behaviour: "A musician is at the harpsichord; he is chatting with his neighbour, he forgets that he is playing a piece of concerted music with others; however, his eyes, his ear, his fingers are not the less in accord with them because of it; not a false note, not a misplaced harmony, not a rest forgotten, not the least fault in time, taste or measure. Now, the conversation ceases, our musician returns to his part, loses his head and does not know where he has got to. If the distraction of the conscious man had continued for a few more minutes, the unconscious animal in him would have played the piece to the end without his having been aware of it."⁸

So the musician, if Diderot is right, sees without being aware of seeing, hears without being aware of hearing. Experimental psychologists have studied similar examples under controlled laboratory conditions and confirmed that the phenomenon is just as Diderot described: while consciousness takes off in one direction, behaviour may sometimes go in quite another. Indeed consciousness may be absent altogether. A sleep-walker, for example, may carry out elaborate actions and even hold a simple conversation without waking up. Stranger things still can happen after brain injury. A person with damage to the visual cortex may lack all visual sensation, be consciously quite blind, and none the less be capable of "guessing" what he would be seeing if he could see.⁹ I have met such a case: a young man who maintained that he could see nothing at all to the left of his nose, and yet could drive a car through busy traffic without knowing how he did it.

So, that is what I am *not* dismissing: the possibility that the brain can carry on at least part of its job without consciousness being present. But what I *am* dismissing is the possibility that when consciousness *is* present it isn't making any difference. And let me now give the two reasons.

First the evolutionary one. When Diderot posed his question, he knew nothing about Darwinian evolution.

He believed in evolution, all right – evolution of the most radical kind: "The plant kingdom might well be and have been the first source of the animal kingdom, and have had its own source in the mineral kingdom; and the latter have originated from universal heterogeneous matter."¹⁰ What is more Diderot had his own theory of selection, based on the

idea of "contradiction": "Contradictory beings are those whose organization does not conform to the rest of the universe. Blind nature, which produces them, exterminates them; she lets only those exist which can co-exist tolerably with the general order."¹¹

Surprising stuff, seeing as it was written in the late 18th century. But note that, compared to the theory Darwin came up with eighty years later, there is something missing. Diderot's is a theory of *extinction*. According to him, the condition for a biological trait surviving is just that it should not contradict the general order, that it should not get in the way. Darwin's theory, on the other hand, is a theory of *adaptation*. According to him, the condition for something's surviving and spreading through the population is much stricter: it is not enough that the trait should simply be non-contradictory or neutral, it must – if it is to become in any way a general trait – be positively beneficial in promoting reproduction.

This may seem a small difference of emphasis, but it is crucial. For it means that when Diderot asked – of consciousness or anything else in nature – "What difference does it make?", he could reasonably answer: "None". But when a modern Darwinian biologist asks it, he cannot. The Darwinian's answer has to be that it has evolved because and only because it is serving some kind of useful biological function.

You may wonder, however: can we still expect consciousness to have a function even if we go along with the idea that it is in fact a "mere surface feature" of the brain? But let's not be misled by the word "mere". We might say that the colours of a peacock's tail were a mere surface feature of the pigments, or that the insulating properties of fur were a mere surface feature of a hairy skin. But it is of course precisely on such surface features that natural selection acts: it is the colour or the warmth that matters to the animal's survival.

Philosophers have sometimes drawn a parallel between consciousness as a surface feature of the brain and wetness as a surface feature of water. Suppose we found an animal made entirely out of water. Its *wetness* would surely be the first thing for which an evolutionary biologist would seek to find a function.

Nonetheless, we do clearly have a problem: and this is to escape from a definition of consciousness that renders it self-evidently useless and irrelevant. Here philosophy of mind has, I think, been less than helpful. Too often we have been offered definitions of consciousness that effectively ham-string the enquiry before it has begun: for example, that consciousness consists in private states of mind of which the subject alone is aware, which can neither be confirmed nor contradicted, and so on. Wittgenstein's words, at the end of his *Tractatus*, have haunted philosophical discussion: "Whereof we cannot speak, thereof we must be silent".

All I can say is that neither biologically nor psychologically does this feel right. Such definitions, at their limit (and they are meant of course to impose limits), would suggest that

statements about consciousness can have no *information content* - technically, that they can do nothing to reduce anyone's uncertainty about what's going on. I find this counter-intuitive and wholly unconvincing. Which brings me to my second reason for dismissing the idea that consciousness is no use to human beings, which is that it is contrary to common sense.

Suppose I am a dentist, and I am uncertain whether the patient in the chair is feeling pain: I ask him "Does it hurt?", and he says "Yes. .. I'm not the kind of guy to show it, but it does *feel* awful". Am I to believe that such an answer – as a description of a conscious state – contains no information? Common sense tells me that when a person describes his states of mind, either to me or to himself (not something he need be able to do, but something which as a matter of fact he often can do), he is making a revealing self-report. If he says, for example, "I'm in pain", or "I'm in love", or "I'm having a green sensation", or "I'm looking forward to my supper", I reckon I actually know more about him; but more important, that *through being conscious* he knows more about himself.

Still, the question remains: what sort of information is this? What is it about? And the difficulty seems to be that whatever it *is* about is, at least in the first place, private and subjective – something going on inside the subject which no one else can have direct access to. I think this difficulty has been greatly overplayed. There is, I'd suggest, an obvious answer to the question of what conscious descriptions are about: namely, that they are descriptions of what is happening inside the subject's *brain*. For sure, such information is "private". But it is private for the good reason that it happens to be his brain, hidden within his skull, and that he is naturally in a position to observe it which the rest of us are not. Privacy is no doubt an issue of great biological and social significance, but I do not see that it is philosophically all that remarkable.

The suggestion that consciousness is a "description of the brain" may none the less seem rather odd. Suppose someone says, for example, "I'm not feeling myself today", that certainly doesn't sound like a description of a brain-state. True enough, it does not *sound* like one; and no doubt I'd have trouble in persuading most people that it was so. Few people, if any, naturally make any connection between mind-states and brain-states. For one thing, almost no one except a brain-scientist is likely to be interested in brains as such (and most people in the world probably don't even know they've got a brain). For another, there is clearly a huge gulf between brain-states, as they are in fact described by brain-scientists, and mind-states as described by conscious human beings, a gulf which is practically – and, some would argue, logically – unbridgeable.

Yet is this really such a problem? Surely we are used to the idea that there can be completely different ways of describing the same thing. Light, for example, can be described either as particles *or* as waves, water can be described either as an aggregation of H₂O

molecules *or* as a wet fluid, Ronald Reagan can be described either as an aging movie-actor *or* as the former President of the United States. The particular description we come up with depends on what measuring techniques we use and what our interests are. In that case, why should not the activity of the brain be described either as the electrical activity of nerve cells *or* as a conscious state of mind, depending on who is doing the describing? One thing is certain, and that is that brain scientists have different *techniques* and different *interests* from ordinary human beings.

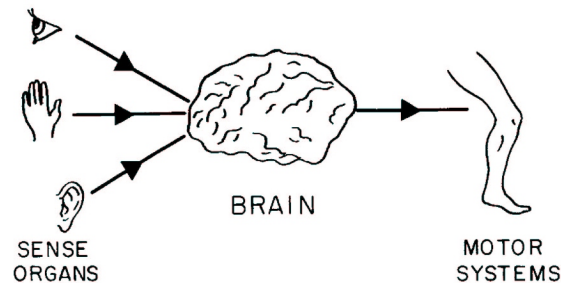
I admit, however, I am guilty of some sleight of hand here. It is all very well to suggest that consciousness is "a description" of the brain's activity, by a subject with appropriate techniques and interests; but what I have not done is to locate this conscious subject anywhere. "To describe" is a transitive verb. It requires a subject as well as an object, and they cannot in principle be one and the same entity. A brain, surely, cannot describe its own activity, any more than a bucket of water can describe itself as wet. In the case of the water, it takes an observer outside the bucket to recognise the water's wetness, and to do so he has to employ certain observational procedures – he has to stick his hand into it, swish it around, watch how it flows.. Who, then, is the observer of the brain?

Oh dear. Are we stuck with an infinite regress? Do we need to postulate another brain to describe the first one, and then another brain to describe that? Diderot would have laughed: "If nature offers us a difficult knot to unravel, do not let us introduce in order to unravel it the hand of a being who then becomes an even more difficult knot to untie than the first one. Ask an Indian why the world stays suspended in space, and he will tell you that it is carried on the back of an elephant .. and the elephant on a tortoise. And what supports the tortoise?.. Confess your ignorance and spare me your elephant and your tortoise."¹²

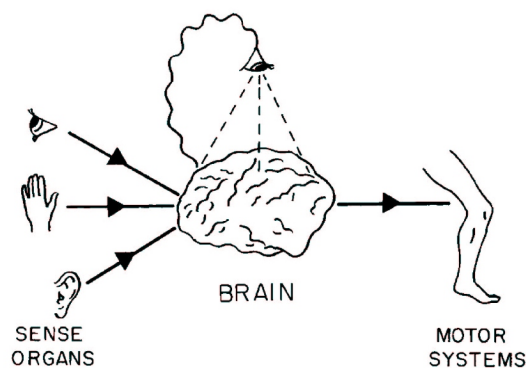
You can hardly expect me, half-way through this essay, to confess my ignorance. And in fact I shall do just the opposite. The problem of self-observation producing an infinite regress is, I think, phony. No one would say that a person cannot use his own eyes to observe his own feet. No one would say, moreover, he cannot use his own eyes, with the aid of a mirror, to observe his own eyes. Then why should anyone say a person cannot, at least in principle, use his own brain to observe his own brain? All that is required is that nature should have given him the equivalent of an *inner mirror* and an *inner eye*. And this, I think, is precisely what she has done. Nature has, in short, given to human beings the remarkable gift of *self-reflexive insight*. I propose to take this metaphor of "insight" seriously. What is more I even propose to draw a picture of it.

Imagine first the situation of an unconscious animal or a machine, which does not possess this faculty of insight, (Figure 4). It has a brain which receives inputs from conventional sense organs and sends outputs to motor systems, and in between runs a highly

sophisticated computer and decision maker. The animal may be highly intelligent and complexly motivated; it is by no means a purely reflex mechanism. But none the less it has no picture of what this brain-computer is doing or how it works. The animal is in effect an unconscious Cartesian automaton.



But now imagine (Figure 5) that a new form of sense organ evolves, an "inner eye", whose field of view is not the outside world but the brain itself, as reflected via this loop. Like other sense organs the inner eye provides a picture of its information field – the brain – which is partial and selective. But equally, like other sense organs, it has been designed by natural



selection so that this picture is a useful one – in current jargon, a "user-friendly" description, designed to tell the subject as much as he requires to know in a form that he is predisposed to understand. Thus it allows him, from a position of extraordinary privilege, to see his own brain-states *as* conscious states of mind. Now every intelligent action is accompanied by the *awareness* of the thought processes involved, every perception by an accompanying sensation, every emotion by a conscious feeling.

Suppose this is what consciousness amounts to. I have written of consciousness as a surface feature of the brain and so I think it is, but you will see now that I am suggesting it is a very special sort of surface feature. For what consciousness actually is, is a feature not of the whole brain but of this added self-reflective loop. Why this particular arrangement should have what we might call the "transcendent" "other-worldly" qualities of consciousness, I do not know. But note that I have allowed here for one curious feature: *the output of the inner eye is part of its own input*. A self-referential system of this sort may well have strange and paradoxical properties – not least that so-called "truth functions" go awry.¹³

Let me recap. We have seen that the brain can do much of its work without consciousness being present; it is fair to assume therefore that consciousness is a second-order property of brains. We have seen that Darwin's theory suggests that consciousness evolved by natural selection; it is fair to assume therefore that consciousness helps its possessor to survive and reproduce. We have seen that commonsense coupled to a bit of self-analysis suggests that consciousness is a source of information, and that this information is very likely about brain-states. So, if I may now make the point that immediately follows, it is fair to assume that access to this kind of second-order information about one's own brain-states helps a person to survive and reproduce.

This looks like progress; and we can relax somewhat. In fact the heavier part of what I have to say is over. You ought, however, to be still feeling thoroughly dissatisfied; and if you are not, you must have missed the point of this whole essay. I set out to ask what difference consciousness makes, and have concluded that through providing insight into the workings of the brain it enhances the chances of biological survival. Fair enough. But the question of course is: How?

The problem is this. We have an idea of what consciousness is doing, namely giving the subject a picture of his own brain activity, but we have no idea yet about what *biological good* this does him in the wider context of his daily life. It is rather as though we had discovered that fur keeps a rabbit warm, but had no idea of why a rabbit should *want* to keep warm. Or, to make a more relevant analogy, it's as though we had discovered that bats have an elaborate system for gathering information about echoes, but had no idea of why they should want such information.

The bat case provides a useful lesson. When Donald Griffin did his pioneering work on echo-location in bats, he did not of course first discover the echo-locating apparatus and then look for a function for it.¹⁴ He began with the natural history of bats. He noted that bats live largely in the dark, and that their whole life-style depends on their apparently mysterious capacity to see without the use of eyes. Hence when Griffin began his investigation of bats' ears and face and brain he knew exactly what he was looking for: a mechanism within the bat

which would allow it to "listen in the dark" – and when he discovered such a mechanism there was of course no problem in deciding what its function was.

This is precisely the tactic we should adopt with consciousness in human beings. Having got this far, we should turn to natural history and ask: is there anything about the specifically human life-style which suggests that people, quite as much as bats, possess a mysterious capacity for understanding their natural environment, for which consciousness could be providing the mechanism.

I shall cut short a long story. When the question is, what would a natural historian notice as being special about the human life-style, I'd say the answer must be this: Human beings are extraordinarily *sociable* creatures. The environment to which they are adapted is before all else the environment of the family, the working group, the clan. Human inter-personal relationships have a depth, a complexity and a biological importance that far exceed those of any other animal. Indeed, without the ability *to understand, predict and manipulate the behaviour* of other members of his own species, a person could hardly survive from day to day.

Now, this being so, it means that every individual has to be, in effect, a "psychologist" just to stay alive, let alone to negotiate the maze of social interactions on which his success at mating and breeding will ultimately rest. Not a psychologist in the ordinary sense, but what I have called a "natural psychologist". Just as a blind bat develops quite naturally the ability to find its way around a cave, so every human being must develop a set of natural skills for penetrating the twilight world of inter-personal psychology – the world of loves, hates, jealousies, a world where so little is revealed on the surface and so much has to be surmised.

But this, when you think about, *is* rather mysterious. Because psychological understanding is immensely difficult; and understanding at the level that most people clearly have it would not, I suspect, be possible at all unless each individual had access to some kind of "black-box" model of the human mind – a way of imagining what might be happening inside another person's head. In short, psychological understanding becomes possible because and only because people naturally conceive of other people as beings *with minds*. They attribute to them mental states – moods, thoughts, sensations and so on – and it is on just this basis that they claim to understand them. "She's *sad* because she *thinks* he doesn't *love* her," "He's *angry* because he *suspects* she's *telling lies*," and so on across the range of human interaction.

I shall not of course pretend that this is news. If it were, it clearly would not be correct. But what we have to ask is where this ordinary, everyday, taken-for-granted psychological model of other human beings originates. How come that people latch on so quickly and apparently so effortlessly to seeing other people in this way? They do so, I

suggest, because that is first of all *the way each individual sees himself*. And why is that first of all the way he sees himself? Because nature has given him an *inner eye*.

So here at last is a worthy function for self-reflexive insight. What consciousness does is to provide human beings with an extraordinarily effective tool for doing natural psychology. Each person can look in his own mind, observe and analyse his own past and present mental states, and on this basis make inspired guesses about the minds of others.

Try it. . . There is a painting by Ilya Repin that hangs in the Tretyakov Gallery in Moscow, its title *They did not expect him*. In slow motion, this is how I myself interpret the human content of the scene:

**** Figure 6. "They did not expect him". ****

A man – still in his coat, dirty boots – enters a drawing room. The maid is apprehensive. She could close the door; but she doesn't – she wants to see how he's received. The grandmother stands, alarmed, as though she's seen a ghost. The younger woman - eyes wide - registers delighted disbelief. The girl – taking her cue from the grown-ups – is suddenly shy. Only the boy shows open pleasure. Who is he? Perhaps the father of the family. They thought he'd been taken away. And now he's walked in, as if from the dead. His mother can't believe it; his wife didn't dare hope; the son was secretly confident that he'd return. Where's he been? The maid's face shows a degree of disapproval; the son's excited pride. The man's eyes, tired and staring, tell of a nightmare from which he himself is only beginning to emerge.

The painting represents, as it happens, a Russian political prisoner, who has been released from the Tsar's jails and come back home. We may not catch the final nuance – more information needed. But try constructing or interpreting a scene like that *without* reference to consciousness, to what *we know* of human feelings – and the depth, its human depth, completely disappears.

I give this example to illustrate just how clever we all are. Consider those psychological concepts we've just "called to mind" – apprehension, disbelief, disapproval, weariness, and so on. They are concepts of such subtlety that I doubt that any of us could explain in words just what they mean. Yet in dissecting this scene – or any other human situation – we wield them with remarkable authority. We do so because we have first experienced their meaning in ourselves.

It works. But I won't hide that there is a problem still of *why* it works. Perhaps we do, as I just said, wield these mental concepts "with remarkable authority". Yet who or what gives

us this authority *to put ourselves in other people's shoes*? By what philosophical license – if there is one – do we trespass so nonchalantly upon the territory of "other minds"?

I am reminded of a story. There was dock strike in London, and enormous lorries were going in and out across the picket lines with impressive notices "By the Authority of H. M. Government," "By the Permission of the Trades Union Congress," "By the Authority of the Ministry of War". Among them appeared a tiny donkey cart, driven by a little old man in a bashed-in bowler-hat, and on the cart was the banner: "By my own bloody authority".¹⁵

That is a good plain answer to the problem. And yet I will not pretend that it will do. Tell a philosopher that ordinary people bridge this gap from self to other "by their own bloody authority", and it will only confirm his worst suspicions that the whole business of natural psychology is flawed. Back will come Wittgenstein's objection that in the matter of mental states, one's own authority is no authority at all: "Suppose that everyone has a box with something in it; we call this thing a 'beetle'. No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle.. it would be quite possible for everyone to have something different in his box .. the box might even be empty."¹⁶

The problem, of course, is not entirely trivial. Strictly speaking, it is true we can never be sure that any of our guesses about the inner life of other people are correct. In a worst case scenario, it is even possible that nature might have played a dreadful trick on us and built every human being according to a different plan. Not just that the phenomenology of inner experience might differ from one person to another, the whole functional meaning of the experience might conceivably be different. Suppose, for example, that when *I* feel pain I do my best to stop it, but that when *you* feel pain you want more of it. In that case my own mental model – as a guide to your behaviour – would be useless.

This worst case scenario is however one which as biologists we can totally discount. For the fact is – it is a biological fact, and philosophers ought sometimes to pay more attention than they do to biology – that human beings are all members of the same biological species: all descended within recent history from common stock, all still sharing more than 99.9% of their genes in common, and all with brains which – at birth at least – could be interchanged without anyone being much the wiser. It is no more likely that two people will differ radically in the way their brains work than that they will differ radically in the way their kidneys work. Indeed in one way it is – if I am right – even less likely. For while it is of no interest to a person to have the same kind of kidney as another person, it *is* of interest to him to have the same kind of mind: otherwise as a natural psychologist he'd be in trouble. Kidney-transplants occur very rarely in nature, but something very much like mind-transplants occur all the time – you and I have just undergone one with those people in the painting. If

the possibility of, shall we call it, "radical mental polymorphism" had ever actually arisen in the course of human evolution, I think we can be sure that it would quickly have been quashed.

So that is the first and simplest reason why this method of doing psychology can work: the fact of the *structural similarity* of human brains. But it is not the only reason, nor in my view the most interesting one. Suppose that all human beings actually had identical brains, so that literally everything a particular individual could know about his own brain would be true of other people's: it could still be that his picture of his own brain would be no help in reading other people's behaviour. Why? Because it might just be the wrong kind of picture: it might be psychologically irrelevant. Suppose that when an individual looks in on his brain he were to discover that the mechanism for speech lies in his left hemisphere, or that his memories are stored as changes in RNA molecules, or that when he sees a red light there's a nerve cell that fires at 100 cps. All of those things would very likely be true of other people too, but how much use would be *this* kind of inner picture as a basis for human understanding?

I want to go back for a moment to my diagram of the inner eye. When I described what I thought the inner eye does I said that it "provides a picture of its information field that has been designed by natural selection to be a useful one – a user-friendly description, designed to tell the subject as much as he requires to know". But at that stage I was vague about what exactly was implied by those crucial words: "useful," "user-friendly," "requires to know". I had to be vague, because the nature of the "user" was still undefined and his specific requirements still unknown. By now, however, we have, I hope, moved on. Indeed I'd suggest we now know exactly the nature of the user. The user of the inner eye is a natural psychologist. His requirement is that he should build up a model of the behaviour of other human beings.

This is where the natural selection of the inner eye has almost certainly been crucial. For we can assume that throughout a long history of evolution all sorts of different ways of describing the brain's activity have in fact been experimented with – including quite possibly a straightforward physiological description in terms of nerve cells, RNA etc. What has happened, however, is that only those descriptions most suited to doing psychology have been preserved. Thus the particular picture of our inner selves that human beings do in fact now have – the picture we know as "us", and cannot imagine being of any different kind – is neither a *necessary* description nor is it *any old* description of the brain: it is the one that has proved most suited to our needs as social beings.

That is why it works. Not only can we count on other people's brains being very much like ours, we can count on the picture we each have of what it's like to have a brain being tailor-made to explain the way that other people actually behave. Consciousness is a socio-biological product – in the best sense of socio and biological.

So, at last, what difference does it make? It makes, I suspect, nothing less than the difference between being a man and a monkey: the difference between we human beings *who know what it is like to be ourselves* and other creatures who essentially have no idea. "One day," Diderot wrote, "it will be shown that consciousness is a characteristic of all beings."¹⁷ I am sorry to say I think that he was wrong. I recognise of course that human beings are not the only social animals on earth; and I recognise that there are many other animals that require at least a primitive ability to do psychology. But how many animals require anything like the level of psychological understanding that we humans have? How many can be said to require, as a biological necessity, a picture of what is happening inside their brains? And if they do not require it, why ever should they have it? What would a frog, or even a cow, lose if it were unable to look in on itself and observe its own mind at work?

I have, I should say, discussed this matter with my dog, and perhaps I can relay to you a version of how our conversation might have gone.

Dog: "Nick, you and your friends seem to be awfully interested in this thing you call *consciousness*. You're always talking about it instead of going for walks."

Nick: "Yes, well it is interesting, don't you think so?"

Dog: "You ask me that! You're not even sure I've got it."

Nick: "That's why it's interesting."

Dog: "Rabbits! Seriously, though, *do* you think I've got it? What could I do to convince you?"

Nick: "Try me."

Dog: "Suppose I stood on my back-legs like a person? Would that convince you?"

Nick: "No."

Dog: "Suppose I did something cleverer. Suppose I beat you at chess."

Nick: "You might be a chess-playing computer. I'm very fond of you, but how do I know you're not just a furry soft automaton?"

Dog: "Don't get personal."

Nick: "I'm not getting personal. Just the opposite in fact."

Dog: (gloomily) "I don't know why I started this conversation. You're just trying to hurt my feelings."

Nick: (startled) "What's that you said?"

Dog: "Nothing. I'm just a soft automaton.. It's all right for you. You don't have to go around *wishing* you were conscious. You don't have to feel *jealous* of other people all the time, in case they've got something that you haven't. . And don't pretend you don't know what it feels like" ..

Nick: "Yes, *I* know what it feels like. The question is do *you*?"

And this, I think, *remains* the question. I need hardly say that dogs, as a matter of fact, do not think (or talk) like this. Do any animals? Yes, there is some evidence that the great apes do: chimpanzees are capable of self-reference to their internal states, and can use what they know to interpret what others may be thinking.¹⁸ Dogs, I suspect, are on the edge of it – although the evidence is not too good. But for the vast majority of other less socially-sophisticated animals not only is there no evidence that they have this kind of conscious insight, there is every reason to think that it would be a waste of time.

For human beings, however, so far from being a waste of time, it was the crucial adaptation - the *sine qua non* of their advancement to the human state. Imagine the biological benefits to the first of our ancestors who developed the capacity to read the minds of others by reading their own – to picture, as if from the inside, what other members of their social group were thinking about and planning to do next. The way was open to a new deal in social relationships, to sympathy, compassion, trust, deviousness, double-crossing, belief and disbelief in others motives.. the very things that make us human.

The way was open to something else that makes us human (and which my dog was quite right to pick up on): an abiding interest in the problem of what consciousness *is* and *why* we have it – sufficient, it seems, to drive biologically normal human beings to sit in a dim hall and listen to a lecture when they could otherwise have been walking in the park.

1. This is an abbreviated version of the James Arthur Memorial Lecture, American Museum of Natural History, New York, 1987. Parts are taken from Nicholas Humphrey, 1986, *The Inner Eye*, London: Faber & Faber.
2. Denis Diderot, 1774-80, *Elements of Physiology*, p. 136, in *Diderot: Interpreter of Nature*, trans. and ed. Jonathan Kemp, London: Lawrence & Wishart, 1937.
3. Denis Diderot, 1774-80, *Elements of Physiology*, p. 139.
4. Charles Darwin, 1838 / 1980, "B Notebook", B232, in *Metaphysics, Materialism and the Evolution of Mind*, ed. Paul H. Barrett, Chicago: University of Chicago Press .
5. Denis Diderot, 1774-80, *Elements of Physiology*, p. 136.
6. J. B. Watson, 1928, *Behaviourism*, London: Routledge & Kegan Paul.
7. Ludwig Wittgenstein, 1958, *Philosophical Investigations*, Oxford: Blackwell.
8. Denis Diderot, 1774-80, *Elements of Physiology*, p. 139.
9. L. Weiskrantz, 1986, *Blindsight*, Oxford: Oxford University Press, Oxford.
10. Denis Diderot, 1774-80, *Elements of Physiology*, p. 136.
11. Denis Diderot, 1774-80, *Elements of Physiology*, p. 134.
12. Denis Diderot, 1747, *Promenade of a Skeptic*, p. 28, in *Diderot: Interpreter of Nature*, trans. and ed. Jonathan Kemp, p. 136, London: Lawrence & Wishart, London.
13. See the brilliant discussion of self reference in Douglas R. Hofstadter, 1979, *Gödel, Escher, Bach*, New York: Basic Books.
14. Donald R. Griffin, 1958, *Listening in the Dark*, New Haven, Ct.: Yale University Press.
15. Cited by A. V. Hill, 1960, *The Ethical Dilemma of Science*, p. 135, New York: Rockefeller Institute.
16. Ludwig Wittgenstein, 1958, *Philosophical Investigations*, I, 294, Oxford: Blackwell.
17. Denis Diderot, 1774-80, *Elements of Physiology*, p. 138.
18. David Premack and Ann Premack, 1983, *The Mind of an Ape*, New York: W. W. Norton.