FIFTY-SEVENTH
JAMES ARTHUR LECTURE ON
THE EVOLUTION OF THE HUMAN BRAIN
1987

THE USES OF CONSCIOUSNESS

NICHOLAS K. HUMPHREY

AMERICAN MUSEUM OF NATURAL HISTORY
NEW YORK : 1987

FIFTY-SEVENTH

JAMES ARTHUR LECTURE ON

THE EVOLUTION OF THE HUMAN BRAIN

FIFTY-SEVENTH
JAMES ARTHUR LECTURE ON
THE EVOLUTION OF THE HUMAN BRAIN
1987

THE USES OF CONSCIOUSNESS

NICHOLAS K. HUMPHREY

AMERICAN MUSEUM OF NATURAL HISTORY
NEW YORK : 1987

JAMES ARTHUR LECTURES ON THE EVOLUTION OF THE HUMAN BRAIN

Frederick Tilney, *The Brain in Relation to Behavior;* March 15, 1932

C. Judson Herrick, *Brains as Instruments of Biological Values;* April 6, 1933

D. M. S. Watson, *The Story of Fossil Brains from Fish to Man;* April 24, 1934

C. U. Ariens Kappers, *Structural Principles in the Nervous System; The Development of the Forebrain in Animals and Prehistoric Human Races;* April 25, 1935

Samuel T. Orton, *The Language Area of the Human Brain and Some of Its Disorders;* May 15, 1936

R. W. Gerard, *Dynamic Neural Patterns;* April 15, 1937

Franz Weidenreich, *The Phylogenetic Development of the Hominid Brain and Its Connection with the Transformation of the Skull;* May 5, 1938

G. Kingsley Noble, *The Neural Basis of Social Behavior of Vertebrates;* May 11, 1939

John F. Fulton, *A Functional Approach to the Evolution of the Primate Brain;* May 2, 1940

Frank A. Beach, *Central Nervous Mechanisms Involved in the Reproductive Behavior of Vertebrates;* May 8, 1941

George Pinkley, *A History of the Human Brain;* May 14, 1942

James W. Papez, *Ancient Landmarks of the Human Brain and Their Origin;* May 27, 1943

James Howard McGregor, *The Brain of Primates;* May 11, 1944

K. S. Lashley, *Neural Correlates of Intellect;* April 30, 1945

Warren S. McCulloch, *Finality and Form in Nervous Activity;* May 2, 1946

S. R. Detwiler, *Structure-Function Correlations in the Developing Nervous System as Studied by Experimental Methods;* May 8, 1947

Tilly Edinger, *The Evolution of the Brain;* May 20, 1948

Donald O. Hebb, *Evolution of Thought and Emotion;* April 20, 1949

Ward Campbell Halstead, *Brain and Intelligence;* April 26, 1950

Harry F. Harlow, *The Brain and Learned Behavior;* May 10, 1951

Clinton N. Woolsey, *Sensory and Motor Systems of the Cerebral Cortex;* May 7, 1952

Alfred S. Romer, *Brain Evolution in the Light of Vertebrate History;* May 21, 1953

Horace W. Magoun, *Regulatory Functions of the Brain Stem;* May 5, 1954

**Fred A. Mettler, *Culture and the Structural Evolution of the Neural System;* April 21, 1955

*Pinckney J. Harman, *Paleoneurologic, Neoneurologic, and Ontogenetic Aspects of Brain Phylogeny;* April 26, 1956

**Davenport Hooker, *Evidence of Prenatal Function of the Central Nervous System in Man;* April 25, 1957

**David P. C. Lloyd, *The Discrete and the Diffuse in Nervous Action;* May 8, 1958

*Charles R. Noback, *The Heritage of the Human Brain;* May 6, 1959

**Ernst Scharrer, *Brain Function and the Evolution of Cerebral Vascularization;* May 26, 1960

Paul I. Yakovlev, *Brain, Body and Behavior. Stereodynamic Organization of the Brain and of the Motility-Experience in Man Envisaged as a Biological Action System;* May 16, 1961

H. K. Hartline, *Principles of Neural Interaction in the Retina;* May 29, 1962

Harry Grundfest, *Specialization and Evolution of Bioelectric Activity;* May 28, 1963

*Roger W. Sperry, *Problems Outstanding in the Evolution of Brain Function;* June 3, 1964

*José M. R. Delgado, *Evolution of Physical Control of the Brain;* May 6, 1965

Seymour S. Kety, *Adaptive Functions and the Biochemistry of the Brain;* May 19, 1966

Dominick P. Purpura, *Ontogenesis of Neuronal Organizations in the Mammalian Brain;* May 25, 1967

*Kenneth D. Roeder, *Three Views of the Nervous System;* April 2, 1968

†Phillip V. Tobias, *Some Aspects of the Fossil Evidence on the Evolution of the Hominid Brain;* April 2, 1969

*Karl H. Pribram, *What Makes Man Human;* April 23, 1970

Walle J. H. Nauta, *A New View of the Evolution of the Cerebral Cortex of Mammals;* May 5, 1971

David H. Hubel, *Organization of the Monkey Visual Cortex;* May 11, 1972

János Szentágothai, *The World of Nerve Nets;* January 16, 1973

*Ralph L. Holloway, *The Role of Human Social Behavior in the Evolution of the Brain;* May 1, 1973

*Elliot S. Valenstein, *Persistent Problems in the Physical Control of the Brain;* May 16, 1974

Marcel Kinsbourne, *Development and Evolution of the Neural Basis of Language;* April 10, 1975

*John Z. Young, *What Squids and Octopuses Tell Us About Brains and Memories;* May 13, 1976

*Berta Scharrer, *An Evolutionary Interpretation of the Phenomenon of Neurosecretion;* April 12, 1977

Lester R. Aronson, *Forebrain Function in Vertebrate Evolution;* April 18, 1978

*Leonard Radinsky, *The Fossil Record of Primate Brain Evolution;* March 26, 1979

Norman Geschwind, *Anatomical Asymmetry of the Brain in Humans and Animals: An Evolutionary Perspective;* April 7, 1980

Irving T. Diamond, *Evolution of the Primate Neocortex;* March 23, 1981

*Robert D. Martin, *Human Brain Evolution in an Ecological Context;* April 27, 1982

Eric Kandel, *Molecular Explorations into Learning and Memory;* April 27, 1983

*Alexander Marshack, *Hierarchical Evolution of the Human Capacity; The Paleolithic Evidence;* May 1, 1984

Yves Coppens, *Environment, Hominid Evolution, and the Evolution of the Brain;* April 16, 1985

Roger A. Gorski, *Sexual Differentiation of the Brain: from Birds to Rats to Man;* April 22, 1986

*Nicholas K. Humphrey, *The Uses of Consciousness;* April 7, 1987

---

*Published versions of these lectures can be obtained from The American Museum of Natural History, Central Park West at 79th St., New York, N.Y. 10024.

**Out of print.

†Published version: *The Brain in Hominid Evolution,* New York: Columbia University Press, 1971.

JAMES ARTHUR
1842–1930

Born in Ireland and brought up in Glasgow, Scotland, James Arthur came to New York in 1871. Trained in mechanics and gear-cutting, he pursued a career in the manufacture and repair of machinery, during the course of which he founded a number of successful businesses and received patents on a variety of mechanical devices. His mechanical interests evolved early into a lifelong passion for horology, the science of measuring time, and he both made some remarkable clocks and assembled an important collection of old and rare timepieces.

Early in this century James Arthur became associated with the American Museum of Natural History, and began to expand his interest in time to evolutionary time, and his interest in mechanisms to that most precise and delicate mechanism of them all, the human brain. The ultimate expression of his fascination with evolution and the brain was James Arthur's bequest to the American Museum permitting the establishment of the James Arthur Lectures on the Evolution of the Human Brain. The first James Arthur Lecture was delivered on March 15, 1932, two years after Mr. Arthur's death, and the series has since continued annually, without interruption.

THE USES OF CONSCIOUSNESS

As is customary on occasions such as this, I should start by thanking my hosts for the surprising gamble they have taken in asking me to give this lecture. The James Arthur Lecture on the Evolution of the Human Brain has brought to this museum a series of distinguished scientists who, in the laboratory or field, have made important discoveries about the workings of the human brain or about human evolution. I have done neither. I hope, therefore, that you will not be too surprised when I tell you that in my lecture this evening I shall have no new research findings to report. In truth I shall have rather little to say about the facts of evolution, and even less to say about the detailed workings of the brain.

I am going to be talking about consciousness—about what consciousness is, and about what part it plays in the natural history of human beings. I shall try to give scientific answers to both questions. But they will, I should say, be "armchair answers": based not so much on new experiments or novel facts, as on a reassessment of facts that we already know. Armchair theorizing has frequently been castigated as too easy. "Unfortunately," Diderot wrote in 1754, "it is easier and quicker to consult oneself than to consult nature."[1] I take his point. But consulting nature on the subject of consciousness is not, I'm afraid, something that anyone that I know is able to do. And since it so happens that one of the few sure facts about consciousness is that every one of us has experienced it in his or her own person, to consult oneself may not be such a bad plan after all.

But that's where *I* must take a gamble. When I talk about consciousness, I'm talking about inner experience—about what it *feels* like to be oneself, to have sensations, thoughts, moods, desires, subjective reasons for one's actions. And if you are to understand—or even be interested—in what I am going to say, I have to assume that we have some sort of shared reference point. But do we? Consciousness is a notoriously tricky concept. No one ever taught us how to use the word correctly. When we were small and learning language, no one ever pointed out an example of consciousness and said, "That's consciousness; remember that next time you see it"—in the way they might have pointed out other things in our envi-

ronment and said, "That's a rabbit . . . or a Spring day . . . or a map of New York." If I and you *do* use the word in the same way, we do not have our nursemaids to thank for it. Perhaps we should thank Mother Nature—well, we shall see.

When I was asked for a title for the lecture, I suggested "The Uses of Consciousness." But, now that it comes to it, I realize I am going to spend a fair bit of time on prior issues—what consciousness is good for as a *concept,* before I get to the question of what it's good for *in our lives.*

A lecture should have a hero. I give you Denis Diderot—the 18th century French philosopher, novelist, aesthetician, 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 the *Elements of Physiology* 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? I shall—if you'll allow me—use the term "soul" and the more modern terms "consciousness" or "self-awareness" more or less interchangeably throughout this talk, without (I hope) misrepresenting anyone or anything. That granted, I have in Diderot's two questions the framework for everything I want to say.

The context for those questions is not hard to guess. Diderot was simultaneously appalled and fascinated by the dualistic philosophy of Réné Descartes. Diderot wrote:

> A tolerably clever man 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]

Fig. 1. Denis Diderot, 1713–84.

Ridiculous it may have been. But sixty 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 realization 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 realization 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.

The philosopher Dan Dennett, for example, is all for programs. There cannot be anything, Dennett maintains, so especially special about nerve cells. If a human brain can carry out the logical functions that, when translated into behavior, persuade us that it's conscious, then so too—at least in theory—could an artificial brain. I say "persuade us that it's conscious" because that—in Dennett's view—is precisely what human brains do: we don't *know* that any other human being is conscious; we are simply led to believe that they are by their behavior.[5] I need hardly tell you that not everyone accepts this way of looking at things. John Searle, for example, finds the idea of artificial consciousness deeply troubling. For him, there is a fundamental distinction to be drawn between a human who is "genuinely" conscious, and a machine which merely behaves "as if" it were conscious.[6] But this is just the distinction that Dennett says is philosophically no good. And so the argument goes on—at the level, so far as I can read it, of "'Tis . . . 'Tisn't."

This is not a dispute on which I want to dwell, partly because I'm not sure that it will ever be resolved, but chiefly because it seems

to me to jump 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 what exactly are our expectations, and how might we account for them? In short, what do we think consciousness *produces*? It brings me directly to Diderot's second question. "If a machine could be united to a soul, what effects—if any—would it have?"

When Diderot asked it, my guess is 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 surely be indistinguishable from a machine without one:

> What difference 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?[7]

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 do not want to pin onto Diderot the authorship of the idea of the functional impotence of souls. But whenever it came, and whether or not Diderot got there, the realization 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 Behaviorist psychologists. There was J. B. Watson, in 1928, arguing that the science of psychology need make no reference to consciousness:

> The behaviorist sweeps aside all medieval conceptions. He drops from his scientific vocabulary all subjective terms such as sensation, perception, image, desire, and even thinking and emotion.[8]

And there, as philosophical backup, was Wittgenstein, arguing that concepts referring to internal states of mind have no place in the "language game."[9] If nothing else, it was an idea to tease one's schoolfriends 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." But I called it a naughty idea, and it is an idea which has had a good run, and now can surely be dismissed.

I will 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. I freely admit that in certain respects the behaviorists may have been right.

Diderot gives a nice example of *unconscious* behavior:

> 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.[10]

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, behavior 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.[11] I have met such a person: 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 reaons.

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.[12]

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.[13]

Surprising stuff, seeing as it was written in the late 18th century. But note that, compared to the theory Darwin came up with 80 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 noncontradictory 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 asks—of consciousness or anything else in nature—"What difference does it make?" he can 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.

Then, either we throw away the idea that consciousness evolved by Darwinian natural selection, or else we have to find a function for it. We can, of course, throw it away. Perhaps Darwin himself did: he was, as I mentioned, prepared to imagine that consciousness has been in some way "super-added"—presumably by some non-natural process. But that is no reason why we should go along with him. I assume—I hope I'm right in this—that everyone present here

is a Darwinian, and that you, like me, would *like* to find a function for consciousness.

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? Well, let's not be misled by the word "mere." We might say that the colors 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 color or the warmth that matters to the animal's survival or reproductive success. 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 that is to escape from a definition of consciousness that renders it self-evidently useless and irrelevant. Here the philosophy of mind has, I think, been less than helpful. Too often we have been offered definitions of consciousness that effectively hamstring the inquiry 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 that 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 that counterintuitive and wholly unconvincing. Which brings me to my second reason for dismissing the idea that consciousness is of no use to human beings, which is that it is contrary to common sense.

Suppose I am a dentist, and 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 state 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 that 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, at least in the first place, is private and subjective—something going on inside the subject which no one else can have direct access to. I think that 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 don't see that it is philosophically all that remarkable.

My suggestion that consciousness is a "description of the brain" may nonetheless 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. I agree. Of course it doesn't *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 that 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 H_2O molecules *or* as a wet fluid, Ronald Reagan can be described either as an aging movie-actor *or* as the 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's doing the describing? One thing is certain: that brain scientists have different *techniques* and different *interests* from ordinary human beings.

I admit, however, that 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 failed to do 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 recognize 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. Am I stuck with an infinite regress? Do I 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.[14]

You can hardly expect me, halfway through this lecture, 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

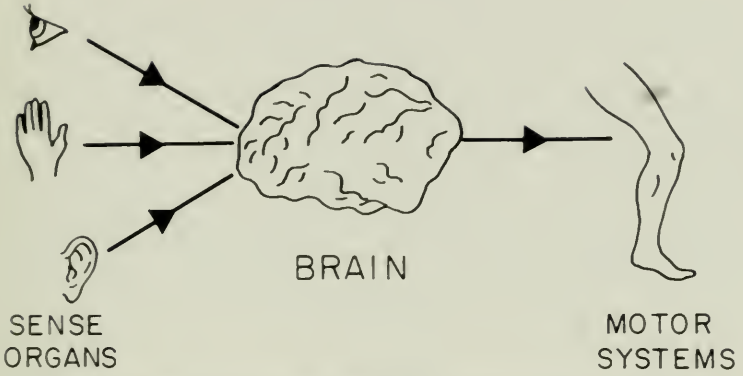

Fig. 2. Diagrammatic representation of an entity lacking insight.

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 that, 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.

I would ask you to imagine first the situation of an unconscious animal or a machine, which does not possess this faculty of insight (fig. 2). 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 decisionmaker. The animal may be highly intelligent and complexly motivated; it is by no means a purely reflex mechanism. But nonetheless 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 (fig. 3) 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

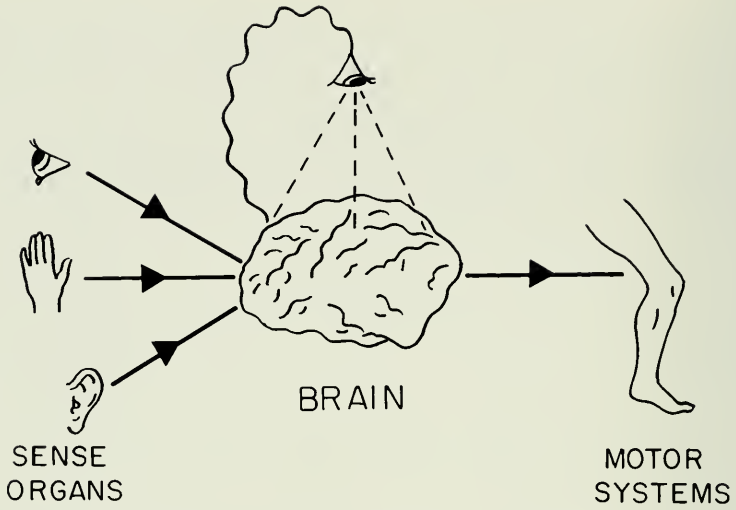

Fig. 3. Diagrammatic representation of an entity possessing insight.

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 talked of consciousness as a surface feature of the brain and so I think it is, but you will see now that I'm suggesting it is a very special sort of surface feature, which has in a sense been "super-added." For consciousness actually 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 I would just note that I have allowed here for one curious feature: *the output of the inner eye is part of its own input.* As I expect you know, a self-referential system of this sort may well have strange and paradoxical properties—not least that so-called "truth functions" go awry.[15]

Let me recap. We've 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've 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've seen that common sense 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.

That looks like progress; and I think we can relax somewhat. In fact, the heavier part of what I have to say is over. I imagine, though, that you are still feeling thoroughly dissatisfied; and if you are not, you must have missed the point of this whole lecture. 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* that does him in the wider context of his daily life. It's rather as though we'd 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 seems to me to provide a useful lesson. When Donald Griffin did his pioneering work on echolocation in bats,[16] he did not, of course, first discover the echo-locating apparatus and then look for a function for it. He began with the natural history of bats. He noted that bats live largely in the dark, and that their whole lifestyle 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.

I think 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 lifestyle 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 (the substance of two books I have written round this issue[17]). When the question is, what would a natural historian notice as being special about the human lifestyle, 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 interpersonal 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 behavior* of other members of his own species, a person could hardly survive from day to day.

Now, that 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 interpersonal 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 that, when you think about it, *is* rather mysterious. For 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's on that 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 ought to ask is where this ordinary, everyday, taken-for-granted psychological model of other human beings originates. Why do 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 there 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 analyze his own past and present mental states, and on that 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* (fig. 4). In slow motion, this is how I myself interpret the human content of the scene:

> 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. Neither you nor I may catch the final nuance—more information needed. But try interpreting a scene like this *without* reference to

Fig. 4. Ilya Repin: "They did not expect him," 1884. Moscow, Tretyakov Gallery.

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 a 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."

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.[18]

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's 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 behavior—would be useless.

This worst-case scenario is, however, one which as biologists we can totally discount. For the fact is—it's 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 percent of the 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'll differ radically in the way their kidneys work. Indeed in one way it is—if I'm 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's 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 behavior. 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 *this* kind of inner picture be as a basis for human understanding?

I want to go back for a moment to my diagram of the inner eye (fig. 3). 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. In the last half hour, 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 behavior of other human beings.

That 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."[19] I am sorry to say I think that he was wrong. I recognize, of course, that human beings are not the only social animals on earth; and I recognize 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 that, 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.[20] 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, far from being a waste of time, it was the crucial adaptation—the sine qua non of our 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.

NOTES

1. Denis Diderot, *On the Interpretation of Nature,* 1754, p. 44. A selection of Diderot's writings is contained in *Diderot: Interpreter of Nature,* ed. Jonathan Kemp. Lawrence & Wishart, London, 1937. All page numbers refer to this edition.
2. *Elements of Physiology,* 1774–80, p. 136.
3. *Elements of Physiology,* p. 139.

4. Charles Darwin, "B Notebook," B232, 1838, in *Metaphysics, Materialism and the Evolution of Mind,* ed. Paul H. Barrett. University of Chicago Press, Chicago, 1980.

5. Dan Dennett, "Evolution, Error and Intentionality," circulating manuscript, 1986.

6. John Searle, *Minds, Brains and Science.* BBC Reith Lectures, London, 1984.

7. *Elements of Physiology,* p. 136.

8. J. B. Watson, *Behaviorism.* Routledge & Kegan Paul, London, 1928.

9. Ludwig Wittgenstein, *Philosophical Investigations.* Blackwell, Oxford, 1958.

10. *Elements of Physiology,* p. 139.

11. L. Weiskrantz, *Blindsight.* Oxford University Press, Oxford, 1986.

12. *Elements of Physiology,* p. 136.

13. *Elements of Physiology,* p. 134.

14. *Promenade of a Sceptic,* 1747, p. 28.

15. Douglas R. Hofstadter, *Godel, Escher, Bach.* Basic Books, New York, 1979.

16. Donald R. Griffin, *Listening in the Dark.* Yale University Press, New Haven, 1958.

17. Nicholas Humphrey, *Consciousness Regained.* Oxford University Press, Oxford, 1983; *The Inner Eye.* Faber and Faber, London, 1986.

18. Ludwig Wittgenstein, *Philosophical Investigations.* Blackwell, Oxford, 1958.

19. *Elements of Physiology,* p. 138.

20. David Premack and Ann Premack, *The Mind of an Ape.* W. W. Norton, New York, 1983.

21. Parts of this lecture have appeared in Nicholas Humphrey, *The Inner Eye,* 1986; *The Guardian,* 28 May, 1986; *Mindwaves,* eds. Colin Blakemore and Susan Greenfield. Blackwell, Oxford, 1987.