SIXTY-SIXTH
JAMES ARTHUR LECTURE ON
THE EVOLUTION OF THE HUMAN BRAIN
1996

DO HORSES GALLOP IN THEIR SLEEP? CONSCIOUSNESS, EVOLUTION, AND THE PROBLEM OF ANIMAL MINDS

MATT CARTMILL

AMERICAN MUSEUM OF NATURAL HISTORY
NEW YORK : 1996

SIXTY-SIXTH

JAMES ARTHUR LECTURE ON

THE EVOLUTION OF THE HUMAN BRAIN

SIXTY-SIXTH
JAMES ARTHUR LECTURE ON
THE EVOLUTION OF THE HUMAN BRAIN
1996

DO HORSES GALLOP IN THEIR SLEEP? CONSCIOUSNESS, EVOLUTION, AND THE PROBLEM OF ANIMAL MINDS

MATT CARTMILL

AMERICAN MUSEUM OF NATURAL HISTORY
NEW YORK : 1996

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

Stephen J. Gould, *Chomsky Under the Spandrels of San Marco;* April 5, 1988

*Harry J. Jerison, *Brain Size and the Evolution of Mind;* October 10, 1989

Paul H. Harvey, *Comparing Brains;* March 20, 1990

Jeffrey T. Laitman, *Evolution of the Vocal Tract and the Origins of Speech;* May 7, 1991

*Dean Falk, *The Evolution of the Human Brain and Cognition in Hominids;* April 14, 1992

Alan Thorne, *A Biological Basis for the Beginnings of Art*? April 26, 1993

Niles Eldredge, *Mind Over Matter: The Evolving Place of Humans in Nature*; April 11, 1994

Este Armstrong, *Expansion and Stasis in Human Brain Evolution: Analyses of the Limbic System, Cortex and Brain Shape*; April 17, 1995

*Matt Cartmill, *Do Horses Gallop in their Sleep? Consciousness, Evolution, and the Problem of Animal Minds*; April 30, 1996

---

*Published versions of these lectures can be obtained from Publications, Dept. of Anthropology, 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.

DO HORSES GALLOP IN THEIR SLEEP? CONSCIOUSNESS, EVOLUTION, AND THE PROBLEM OF ANIMAL MINDS

It is an honor to be asked to deliver the annual James Arthur Lecture on the Evolution of the Human Brain. The honor is especially great for me, because I'm not an expert on the evolution of the brain. I want to talk to you this evening not so much about brain evolution—which some of you know more about than I do—but rather about the evolution of the mind: in particular, that aspect of our mental lives that we call awareness or consciousness.

I've chosen to talk about consciousness for two reasons. For one thing, nobody understands much about it, and so I'm almost as qualified to talk to you about it as anybody is. More importantly, the appearance of consciousness is what gives the whole subject of brain evolution its importance. Why do I say this? Because consciousness is the only thing that gives importance to anything at all.

All values in the world have their source in the objective fact that the world contains subjective consciousness. That may sound like metaphysics, but it's really just a matter of common sense. To put it another way: life isn't worth anything if you spend it in a coma.

Let me propose a thought experiment. I would like you to imagine three things. First, imagine that we've discovered that a certain clump of nerve cells in the brain is essential for conscious awareness. Second, imagine that a certain drug suppresses neural activity in just this nucleus, with no effect on the rest of the brain. Subjects who take this drug do things as usual, but they experience nothing until the drug wears off. The drug converts them into sleepwalkers. Finally, imagine that I've developed a new form of this drug, which has *permanent* effects. It abolishes consciousness forever, with no effect on behavior. I want to test it on you. How much will you charge to take it?

I see no volunteers, and I think the question answers itself. Spending your life as a sleepwalker is equivalent to being dead; and so you will charge me whatever price you would charge to commit suicide. Some people might accept the deal, but only if their lives were so desperate that oblivion seemed attractive.

Oblivion attracts the wretched because the absence of consciousness erases all values, negative as well as positive. To the unconscious, nothing is either a blessing or a curse, any more than it is to an inanimate object. You are not being cruel to your car when you fail to service it, or kind when you rotate its tires. Nothing is either a benefit or an injury from a car's standpoint, because it has no standpoint. Having no awareness, it has no interests. And the same, I think, is true of other permanently unconscious things, both organic and inorganic. For example, a species has no interest in surviving, because it isn't the sort of thing that can have interests. When we speak of an extinct species as an evolutionary failure, we're being anthropomorphic. You can't fail without intentions; and intentions are impossible in the absence of consciousness. Even in the case of *human* life, our legal codes make similar judgments about hospital patients whose brains have gone electrically silent. These people are human, and alive; but because we believe them to be permanently unconscious, we deem them to have no further interests, and so we demote them to a purely instrumental value—for example, as a source for organ transplants.

These facts are fairly obvious, but I want to stress them here at the start of these remarks to dispel the notion that conscious awareness is too metaphysical and subjective a phenomenon for science to concern itself with. As the source of all value in our lives, it should be at the top of the scientific agenda. Yet in spite of its fundamental importance, consciousness is a subject that most scientists are reluctant to deal with. We know practically nothing about either its mechanisms or its evolution.

In fact, many distinguished scientists and philosophers believe that consciousness has no evolutionary history, because they think human beings are the only creatures that have it. Even many of those who suspect that some other animals may be conscious doubt that we can ever know for sure, and therefore would prefer to exclude this whole subject from the scientific world picture. Most scientists, I think, will admit in private that our close animal relatives probably have mental lives something like ours, because after all they have bodies and brains and behavior that resemble ours more or less closely. But a lot of scientists are reluctant to say so plainly and publicly; and those that do can count on being accused of sentimentality and anthropomorphism.

I ran into this recently when I wrote a rejoinder to an opinion piece that appeared in a major biology journal. Its author had condemned animal-rights activists for failing to understand that an animal's major purpose in Nature is to be eaten by others. In replying, I asked what could possibly be meant by talking about an animal's purpose in Nature. Nature isn't the sort of thing that has purposes or intentions. Nature is just The Way Things Are. Only certain animals have purposes and intentions—and they never include being eaten by others. When the journal's editor read this, he at once demanded that I cite some published studies to support my dubious claim that *animals* can have intentions. So I smiled and changed the sentence to read, "Some animals (for example, human beings) have intentions." There were no more objections.

People have in fact done experimental studies to test the proposition that nonhuman animals sometimes have intentions (Heyes and Dickinson, 1993). But as far as I know, nobody has felt the need to run experiments to determine whether you and I have intentions. We all know that *we* ourselves have them; and we know that other people are built like us, and behave like us, and act as though they have intentions. That's all we need to know. Yet somehow the same sort of evidence doesn't settle the issue when other species are in question.

Animal intentionality, or what looks a lot like it, is of course a commonplace everyday phenomenon. For instance, most of us who own dogs have probably had the experience of seeing our dog search out a favorite toy and bring it to us in the hopes of getting us to play with him. It's difficult and awkward even to describe these familiar experiences without saying things like, "The dog was trying to find his ball," or "The dog wanted me to play with him." But scientists aren't supposed to say things like that, at least when we have our lab coats on. If we discuss such things at all, we prefer to do so in some way that doesn't involve attributing intentions or any other mental states to the dog.

There are at least two ways we can do this. First, we can use clumsy behavioral circumlocutions for mental language. Instead of saying, "The dog looked for his ball until he found it," we can say something like, "The dog exhibited repeated bouts of investigative behavior, which ceased after he contacted the ball." This somehow manages to suggest that the dog wasn't thinking about the ball while he was looking for it, and that he didn't perceive anything when he got it in his mouth.

Second, if we find these circumlocutions silly and tedious, we can adopt some variant of what is sometimes called "logical behaviorism," in which the mental words are still used but are redefined in terms of the probabilities of various behaviors. In this view, a dog's intentions and desires and beliefs turn out, when properly understood, not to be something inside the dog, but theoretical constructs pinned on the dog by a human observer. Therefore, the human observer can know whether the dog has intentions and desires and beliefs, but the dog can't. The philosopher Daniel Dennett is probably the best-known advocate of this position at the moment. In Dennett's reading, dogs have real beliefs and intentions—but so do computers and thermostats and alarm clocks, because believing something is literally identical with *behaving* as though you believed it. The mental states that accompany your behavior are irrelevant; and in Dennett's view, they're unique to human beings anyway (Dennett, 1987a, 1987b, 1991, 1995).

Why do scientists and philosophers go through all these contortions to avoid attributing mental states to animals? There are several reasons. Some of them are better than others.

Let's start with the best ones. There's no doubt that sentimentality and uncritical anthropomorphism are real temptations, and that they should be avoided in describing and analyzing the behavior of nonhuman organisms. A lot of us succumb to these temptations. We all know people who insist on telling you what kind of music their begonia likes or what their cat thinks about Rush Limbaugh. These people are mistaken. And scientists sometimes make similar mistakes. Some of the early Darwinians in particular were guilty of this sort of thing. Because Darwin's opponents often cited the mental and moral differences between people and beasts as reasons for rejecting the whole idea of evolution, many of his early followers tried to play down those differences by repeating anecdotes they had heard about the nobility of dogs, the cunning of mules, and the self-sacrifice of chickens.

The British psychologist C. Lloyd Morgan was dismayed by the early Darwinians' uncritical attribution of human mental states to animals, and he tried to put a stop to it. In 1894, Morgan laid down the following law in his book, *An Introduction to Comparative Psychology*:

> —*In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale.*

Successive generations of experimental psychologists have adopted this dictum as a fundamental axiom called *Morgan's Canon*. It's generally thought of as a special case of Occam's Razor, the principle that you shouldn't make up entities unless you have to. For instance, you shouldn't posit a mysterious life force in living things if you can explain all the phenomena of life in terms of chemistry. Likewise, if you can explain an animal's behavior as, say, a conditioned reflex, you shouldn't try to interpret it as the outcome of volition or thinking. By this view, we are required to deny mental events in animals whenever we can, in the name of parsimony.

All of this sounds reasonable, but there's a fundamental flaw in it. Because *we* have mental events, we already know that there *are* such things in the universe. Denying them to animals therefore doesn't *save* anything; we have the same number of entities on our hands no matter what we decide about animal minds. Occam's Razor doesn't provide any support for Morgan's Canon. In fact, some animal-rights philosophers (e.g., Regan, 1983: 29) claim that Occam's Razor is on *their* side. They argue that if we're going to invoke intentions, desires, beliefs, and other mental phenomena in accounting for our own actions, we should explain other animals' behavior in similar terms whenever we can—again, in the name of parsimony.

Morgan himself agreed that it would be simpler to assume that other animals have mental lives like ours; but he insisted that simplicity is no guide to truth. He felt that his Canon was justified not by the principle of parsimony, but by the theory of evolution. His argument to that effect (Morgan, 1894: 55–59) involved a very 19thcentury picture of evolution as the grand story of our ancestors' climb from the primordial ooze up to the human condition, which furnishes the standard by which "psychical faculties" are to be judged as "higher" or "lower in the psychological scale." "Higher" here turns out to mean "distinctively human," and "lower" means "shared with other species." So the true, underlying meaning of Morgan's Canon is something like this:

> *—In no case may we interpret an animal's actions as the outcome of humanlike mental events, if we can find any other way of explaining them.*

This still sounds like a prudent proposition. But why is it? Why should we assume a priori that if *we* have something, then animals don't? What risks, exactly, are we guarding against here? Why is it *safer* to assume that human properties are unique? Why wouldn't it be a safer bet to assume the opposite?

The problem with Morgan's Canon comes into sharp focus if we transfer the argument from the brain to the kidney. Consider this version:

> *In no case may we interpret an animal's urine as the outcome of humanlike biochemical processes, if we can find any other way of explaining it.*

If Morgan's Canon represents a safe assumption, so does this one. But it's obvious that this version is ridiculous, and that physiologists would think I was crazy if I insisted that they adopt this rule to avoid the temptations of anthropomorphism. Then why does Morgan's Canon *seem* so much more plausible than this one? Are neurologists just more gullible than urologists? Or is there something special about events in the brain that makes them different from events in the kidneys?

Part of the answer, of course, is that we don't care about kidneys the way we care about brains, because brain events are a source of human status and kidney events are not. Our mental abilities are markers of the boundary between animals and people, which is one of the two primary lines that we use to divide up the moral universe. Because nonhuman animals lack some of our mental abilities, we regard them as property, to be used for our ends in any way we choose—on the dinner table, or in scientific experiments, or transformed into soap and shoes and lampshades. The only moral constraint that we observe on our use of other animals is an obligation not to make them suffer. And we acknowledge *that* duty only because we believe that at least some of the animals are on our side of the *second* big line we draw across the moral landscape—the boundary between sentience and nonsentience, between things that are conscious and things that aren't. So both of our major moral boundaries are defined by things that go on in the brain.

Brain events, then, have moral implications that kidney events don't, which is why we're so generous about recognizing humanlike urine in other animals and so stingy about recognizing humanlike behavior. I want to point out in passing that the whole notion of "behavior" hinges on this moral aspect of neurology. When people like Stephen Gould argue that human behavior isn't biologically determined (e.g., Gould, 1981: 327), they aren't thinking about such body movements as the contractions of the heart or the intestines. Those movements are quite thoroughly determined biologically; but we don't think of them as *behavior,* because we don't regard ourselves as *responsible* for them. Having a spastic colon is not bad behavior. "Behavior" means *voluntary* movements—movements produced by striated muscle under conscious cortical control. If a movement is produced by cardiac or smooth muscle, or by striated muscle under the exclusive control of the brainstem or spinal cord, we don't call it behavior; we call it physiology. The distinction we draw between physiology and behavior is a projection of our concept of moral agency—which ultimately depends on the fact of consciousness. Unconscious actions are by definition involuntary.

Up to this point, I have been assuming that mental events are, or are produced by, events in the brain. Scientists rarely question this assumption. However, philosophers question it a lot. Some of them argue that mental events can't be equated with brain events, because we can see other people's brains but not their minds. Brain events, they point out, are objective and public; mental events are subjective and private. This is the other crucial difference between the brain and the kidneys—and the other source of scientists' uneasiness about the question of animal consciousness.

The intrinsic *subjectivity* of consciousness makes scientists uncomfortable. Being conscious is the same thing as having private experiences; and the scientific method is fundamentally committed to the assumption that private experiences don't count as evidence. Only publicly accessible and repeatable experiences have that status. If somebody makes a claim that you can't check out for yourself, then you're not obliged to take it seriously. This makes science constitutionally anti-authoritarian, which is good; but it also makes it unreceptive to claims about consciousness and its contents. Most of the recent literature on the subject of consciousness is not really about consciousness at all, but about either neurology or behavior. These are public phenomena, and scientists know how to deal with them. So they spend a lot of time trying to convince themselves that studying these things is somehow the same thing as studying consciousness—like the drunken man in the story who lost his wallet in Central Park but went looking for it in Times Square because the light was better.

The field of computer science called artificial intelligence grew out of these assumptions. In a classic paper published in 1950, the English computer theorist Alan Turing offered a test for telling whether machines can think. He called it "the imitation game." Suppose, he said, that we can write a program that will exchange e-mail with you. If, after five minutes of sending messages back and forth, you can't tell whether you've been chatting with a human being or a computer, then the machine has a human mind—because that's what having a human mind means: being able to carry on a human conversation. What other test could there be? And Turing (1950) predicted that some of us would see such machines within our lifetimes. Here's the quote:

> I believe that in about fifty years' time it will be possible to program computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than a 70 per cent chance of making the right identification after five minutes of questioning.

Check out those numbers. It's about fifty years later, and 10^9 equals around 16 megabytes. You can buy the supercomputer of Alan Turing's fondest dreams off the shelf at Sears for the price of a beat-up used car; and far bigger machines can be had at higher prices. But none of them has yet been programmed to play the imitation game successfully. What went wrong?

I think what went wrong wasn't just Alan Turing but the whole Western conception of what it means to be human. Our traditions encourage us to define ourselves not by what we *are*, but by how we are *different*: to think of the human essence not in terms of our *properties*, but in terms of our *peculiarities*—the small subset of human traits that we don't share with any other creatures. Many of these human peculiarities hinge on our unique skill in manipulating symbols; and that also happens to be what philosophers get paid for doing. It's not surprising, therefore, that philosophers and professors from Plato on down to Noam Chomsky have told us that juggling words and numbers is the defining excellence that makes people special, and that animals that lack it are mere objects. Marcus Aurelius summed it up in this maxim (*Meditations*, 6.23): "Use animals and other things and objects freely; but behave in a social spirit toward human beings, because they can reason."

Many Western thinkers have gone further and insisted that because animals can't talk, their mental lives are defective in big ways, or even nonexistent. "Thinking," wrote Wittgenstein (1958: 6), "is essentially the activity of operating with signs." That view of thinking naturally appeals to college professors, who sometimes get so consumed by operating with signs that they wander around their campuses talking to themselves and tripping over shrubs. And because nonhuman animals aren't very good at operating with signs, many professorial types have been reluctant to grant that beasts can have mental lives at all. Others suspect that thinking also has something to do with not tripping over shrubs, and that dogs and cats and horses may be as good at it in most ways as a lot of college professors are—and infinitely better at it than any computer is.

Some people argue that since animals don't have words for things, they don't have *concepts*; they can't judge and classify sensations, and therefore they don't really perceive objects. All they do is respond to stimuli. A surprising number of Western philosophers and psychologists, from Augustine (*City of God*, 12.4) on down, have bought into this notion, but it seems to me to be demonstrably false. Many, and maybe most, of our concepts have no words attached to

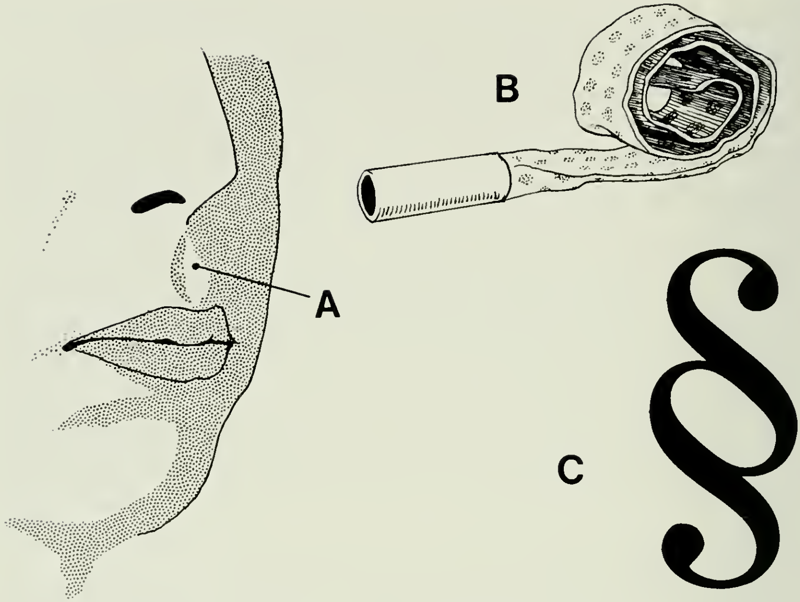

Fig. 1. Three examples of nonverbal concepts: (**A**) an anatomical feature, (**B**) an artifact, and (**C**) a written symbol.

them. Three examples—one natural object, one type of artifact, and one written symbol—are illustrated in figure 1. I suppose there are names for these things, but the only one *I* know is the first one; and I even had to look *that* one up—and I'm an anatomy professor. Yet we all recognize these things as representatives of familiar classes. And if *we* don't need words to have concepts, then neither do dogs, horses, and pigeons.

Because Western thinkers have always attached so much importance to juggling symbols as a marker of human status, and so little importance to walking around without tripping over things (which couldn't be very important, because a donkey can do it just as well as a philosopher), it was inevitable that when we managed to build a symbol-juggling engine—a machine that could beat us all at chess and prove the four-color theorem—our philosophers would try to persuade us that it was human. Once we taught it to play the imitation game, they assured us, it would be just like one of us. But so far, it has proved impossible to program such an engine to succeed at the imitation game. The reason is that, although a computer has many of the symbol-manipulating abilities that we prize so highly, it lacks the subtler and more mysterious skills that come with being a sentient animal, inhabiting and experiencing the world in a living body.

Computer metaphors have come to dominate our thinking about brain processes and mental events. They predispose us to believe that mental events are *algorithmic*—that is, that they are produced by executing a programmatic list of logically connected instructions—and that digital computers (which are algorithm machines) will eventually become conscious if only we can run the right program on the right kind of hardware with the proper stored data. But as the philosopher John Searle (1992) has argued forcefully, there are good reasons for thinking that conscious awareness isn't, and can't be, produced by running a computer program.

A digital computer is essentially a grid of slots, each of which can be either full or empty. We think of these as ones and zeroes. Some of these slots are linked causally by rules of operation, which provide that when a certain pattern shows up in *this* area, the contents of other slots are changed in various ways, which may depend on the contents of yet *other* slots. In modern computers, the ones and zeroes are represented by electrical charges in semiconductors, but they could be represented by anything: holes punched in cards, or beads on wires, or eggs in egg cartons. The medium doesn't matter: what's important is the algorithm. All the operations that you do on a computer could be done in exactly the same way by giving a team of people written instructions for moving eggs around in a football field full of egg cartons, though of course it would take longer. (By the way, a football field full of egg cartons has about 1 megabyte of RAM.)

This fact poses problems for computational theories of the mind. If moving electrical charges around in a certain pattern can produce subjective awareness and bring a mind into existence, then so can moving around a collection of eggs in the same pattern; and if I knew how many eggs to use and what rules of operation to use in moving them, then I could make my egg collection think it was

Elizabeth Dole or the Wizard of Oz. I could get the same effects by making chalk marks on a blackboard, or waving semaphore flags, or singing songs, or tap dancing. All these processes can be computationally equivalent, with algorithms that correspond in every detail; but none of them seems like a plausible way of producing a subjective awareness. And since a digital computer is just another way of instantiating an algorithm, it seems impossible for such a device to become conscious. If we ever succeed in creating an artificial intelligence, it's going to have to be something more than an algorithm machine.

If consciousness isn't algorithmic, then how is it produced? We don't know. The machineries of consciousness are an almost perfect mystery. Neuroscientists and computer scientists have produced a lot of useful and suggestive models of how the brains of animals process sensory data and judge and discriminate among stimuli. We know that such mechanisms exist in our own brains, and that we need them to perceive the world. But although these perceptual mechanisms are *necessary* for consciousness, they aren't *sufficient*, because we can perceive things and respond to them without being aware of them. The unconscious mind is a real phenomenon. Freud's picture of it may have been wrong in its details, but he got the big picture right: most of our mental activity is carried out by subsidiary parts of the mind to which we don't have any direct conscious access.

One much-studied example of this is the phenomenon of blindsight. People with brain injuries to the visual cortex at the back of the head often go blind; they report seeing nothing, they can't read, they walk into things. But if you present things to their eyes and ask them just for fun to guess what they would be seeing if they *could* see, they guess with surprising accuracy. What seems to be happening is that visual centers in the older, subcortical parts of the brain are receiving and processing retinal input, and then forwarding it to the cortical speech centers along pathways that bypass the visual cortex—which are therefore not perceived *as vision* (Weiskrantz, 1986).

The reverse phenomenon, called perceptual defense, is more familiar. People confronted with unwelcome sights or sounds often don't perceive them consciously, though you can show by monitoring their blood pressure or pupil dilation or other responses that they have in fact perceived them and interpreted them correctly. For example, if you present pairs of spoken words to people through stereo headphones and ask them whether they heard the same word in both ears, they fail to hear the difference between disparate pairs significantly more often if one of the two words is an obscenity. At least that used to be true when I ran that experiment in my undergraduate psych lab, though I imagine the perceptual threshhold for tabooed words is lower now than it was back then.

Not only do we perceive many things unconsciously; we can act on those perceptions without being aware of our actions. The most spectacular example of this is sleepwalking.

Sleep takes several forms. In the living brain, waves of nerve-cell discharges travel across the surface of the cerebral cortex like the network of ripples on a swimming pool. We can detect, monitor, and record them as changes in electrical potential on the surface of the overlying scalp. Like the ripples on a pool, these brain waves vary in frequency and amplitude, from big and slow to fine and choppy. They're fine and choppy when we're awake; this is called the alpha rhythm. As we sink deeper and deeper into sleep, the waves become slower, bigger, and more synchronized, because the neurons in one area are all tending to fire at the same time. The waves become slowest and biggest in the so-called delta rhythms of deep sleep, in Stages III and IV. It's hard to wake people up from these stages. In normal awakening, the sleeping brain climbs back up this staircase, through all the successive stages of sleep, back to the alpha rhythms seen in both Stage I sleep and waking consciousness (Hobson, 1989).

Now, here's the strange part. Many people—as many as 30 percent of all children and 7 percent of adults—sometimes get up and start walking around during the deepest, most unconscious part of sleep, in Stage IV. Typically, sleepwalkers open their eyes, sit up in bed with a blank facial expression, pluck aimlessly at the bedclothes, and then rise up and walk. They ignore objects and people nearby, but they usually manage to get around without bumping into things. They may do very complicated and distinctively human things—

talk, make phone calls, get in a car and drive off, or even play musical instruments. Conversely, they may also do very dangerous and stupid things, like walking through glass doors or over cliffs. If you try to wake them up, they struggle violently to get away from you; and if you succeed in awakening them, they're totally confused and have no recollection of what they were doing or how they got there (Hartmann, 1983; Rauch and Stern, 1986; Reite et al., 1990; Thorpy, 1990).

The phenomenon of sleepwalking shows that you can get surprisingly complicated and even distinctively human behavior without consciousness. Some sleepwalkers could pass the Turing test 30% of the time (which is all that Turing demanded) with no difficulty. I know this because the one time I myself encountered a sleepwalker, it took me several minutes to recognize that she was unconscious; and I probably wouldn't have caught on at all if she hadn't been a family member whose behavior I knew very well. All this makes it much harder for us to find out anything about animal awareness. How do we know that animals aren't simply sleepwalking all the time, even when they appear to be awake? Do wolves hunt and horses gallop in their sleep, in the same way that a human somnambulist gets into a car and drives off on the freeway at 65 miles an hour? When the cock crows in the morning, is the farmer the only animal on the farm that wakes up? And if we can do so many things without being conscious, then why did consciousness evolve?

Let's start with that last question first. It's been proposed that consciousness permits you to construct objects in your mind out of the diverse input from several different senses (Jerison, 1973). When we see a car drive past, we don't separately hear its motor *and* see its body *and* smell its exhaust. We perceive one thing—a car passing by. We attribute the sight, sound, and smell to it as its properties; and we get out of the way to avoid the tactile sensations that we expect to go with them.

Neurobiologists refer to this as the "binding" phenomenon. Most animals pretty clearly don't have it. They don't exhibit what's called *cross-modal perception*; that is, they don't recognize an object through one sense if they've experienced it only through another.

They have separate and very mechanical responses to the inputs from different sense receptors. When a frog strikes at a fly, it doesn't see the fly as an object; it's just built to snap at any moving object overhead of a certain size (Barlow, 1953; Barlow et al., 1972). A frog will starve to death in the midst of a heap of freshly killed flies. It doesn't recognize them as flies, and it won't sniff or peck at them or try one to see if it tastes good, the way a mammal or a bird would. It seems reasonable to conclude that frogs aren't conscious. It also seems reasonable to think that frogs might be better off, other things being equal, if they *were* conscious and could perceive flies and other objects.

It has accordingly been suggested that binding and consciousness are different words for the same thing, and that consciousness is adaptive because it allows us to construct objects in our minds, benefit from cross-modal learning, and develop an internal map of the world that lets us anticipate what's going to happen and stay one jump ahead of things—instead of just producing knee-jerk responses to stimuli. If all this were true, then we could tell which animals were conscious by just testing them for cross-modal perception. And this does seem like a valid negative test; that is, animals like frogs, that *don't* have cross-modal perception, probably aren't conscious in any recognizable sense.

But the test won't work the other way around, as evidence *for* consciousness, because people can integrate different sensory modes while they're asleep. For example, sleepwalkers have been known to sit down and play the piano; and anybody who can do that must be putting hearing, touch, and proprioception together into a single experiential construct. If people can do this sort of thing in Stage IV sleep, then binding can take place without conscious awareness.

Then what does consciousness do for us? Well, why should we assume that it does anything? Some people have argued that it confers no adaptive advantage whatever—that it's just an incidental side effect of neurological complexity. But I think that idea can be rejected for Darwinian reasons. If consciousness were a useless epiphenomenon, natural selection would have operated to get rid of it somehow, since we have to pay a high price to have it.

The price we pay for consciousness is unconsciousness, of the special and peculiar sort we call sleep. Mammalian sleep is a complicated and dangerous performance, and most animals don't do it. Invertebrates and cold-blooded vertebrates usually have daily periods of torpor when they hide and rest, but most of them show little or no correlated change in neural activity (Hartse, 1989). Among vertebrates, true sleep, involving a shift from fast to slow waves in the forebrain, appears to be limited to mammals and birds, though there are hints of it in a rudimentary form in some reptiles.

In birds and therian mammals (Hobson, 1989; Amlaner and Ball, 1989; Zepelin, 1994; Siegel et al., 1996), slow-wave sleep is interrupted at intervals by a second kind of sleep called REM (rapid eye movement) sleep. In REM sleep, the EEG goes back to the alpha rhythm. There are synchronized bursts of activity in different sensory areas of the cortex, as well as in the muscles of the eye, the middle ear, and the pinna. The eyes swing this way and that in coordinated tracking movements, in phase with the bursts of nerve-cell activity in the visual cortex; but they do so behind closed eyelids. In short, the brain appears to be seeing and hearing things that aren't there. In the human brain, at any rate, it's doing just that—because REM sleep is associated with dreaming. Another thing that happens during REM sleep is that the body's muscles lose their tone; we go totally limp from the neck down. This happens because nuclei in the brainstem spread chemicals around that raise the transmission threshhold at the synapses linking the brain to the spinal cord. Sensory impulses coming in from the skin have a hard time getting through to the brain, and motor commands coming out of the brain don't get passed along to the spinal cord. In effect, the volume is turned down on the brain-body connections in both directions, so that the dreaming brain can attend to its own fantasies without responding to the world or jerking the body around. When these inhibitory mechanisms don't work in human beings, you get a pathological sort of sleepwalking called REM behavior disorder, in which sleepers act out their dreams—often with traumatic or even fatal results (Mahowald and Schenck, 1989).

Why do we sleep? On the face of it, it sounds like a bad idea to spend about a third of the day plunged into a limp, helpless trance state that leaves you unable to detect or react to danger. Mammalian sleep is so dangerous, complicated, and time-consuming a performance that we feel sure it must have a payoff of some sort, but it's not really clear exactly what it is. Some argue that sleep serves to conserve energy, which is why we see it only in warm-blooded animals. The trouble with this theory is that mammalian sleep uses almost as much energy as wakeful resting. During 8 hours of sleep, a human being saves only about 120 calories (Zepelin, 1994)—the equivalent of a small glass of 2% milk, or three-quarters of a plain bagel. These savings don't seem worth spending a third of your life dead to the world. Another theory holds that sleep is a defense against predators; it's nature's way of telling us to hide when we don't need to be active (Webb, 1974). The problem with this story is that animals that don't sleep also find holes to hide in when they rest; and birds and mammals that are too big to hide in holes or climb trees still have to flop down and fall asleep every day, right out there on the prairie, exposed to every predator in the world. They do it as little as possible—a horse sleeps only about 3 hours a day, of which only some 35 minutes is spent lying down in REM-sleep atonia (Zepelin, 1994)—but they'd be better off if they didn't do it at all. They do it because they have to do it, not to save energy or avoid predators.

So sleep appears to be something imposed upon us, not by our environmental circumstances, but by the needs of the brain itself. Consciousness depletes or damages something in the waking brain, and we can't keep it up indefinitely. If we're forced to stay conscious around the clock, day after day, with no sleep, we soon start manifesting pathological symptoms, beginning with irritability and proceeding through fainting and hallucinations to metabolic collapse and death. And we seem to have a separate need for REM sleep in particular. If you keep waking people up whenever they enter the REM stage of the sleep cycle, they eventually start dropping straight into REM sleep without any slow-wave preliminaries. When this happens spontaneously, it results in a potentially serious behavioral disorder called narcolepsy.

So what is it that sleep does for the mammalian brain? Several people, including Francis Crick (Crick and Mitchison, 1983), have suggested that birds and mammals need to sleep because their behavior is flexible and based on learning and experience, instead of being just a collection of stimulus-response reflexes. This theory holds that behavioral flexibility—free will, if you like—introduces noise into the system and tends to mess up the innate, "hard-wired" responses and behaviors that these animals still need for survival. According to this model, sleep in general—and REM sleep in particular—acts every day to erase the neural irrelevancies, reset all the innate systems, and put everything back in working order, like rebooting a computer.

One fact that supports this model of sleep is the phenomenon called retrograde amnesia. When we're awake, the things that drop out of short-term memory drop out more or less in the order they happened, so that it's easier to remember things that happened two minutes ago than those that happened two hours ago. But when we drowse off, we start forgetting backwards, so that the last things learned are the first to be forgotten; and the longer we sleep, the further back in time the erasing of memory extends (Guilleminault and Dement, 1977). This shows that there is an active process of erasure that is peculiar to the process of sleep.

Another piece of evidence in favor of this model is that we get less and less sleep in general, and REM sleep in particular, as we age. In typical infant mammals, including human babies, over 80% of the sleep cycle is spent in REM sleep. As we reach adulthood, the world becomes more familiar and our behaviors more habitual; our brains get more canalized, new learning becomes less frequent and more difficult—and REM sleep drops to about 20% of the total cycle. Perhaps it does so because there is less new learning to be cleaned up after. It seems significant in this context that highly precocial infant mammals like guinea pigs and ungulates, which pop out of the uterus bright-eyed, bushy-tailed, and ready to start running, have low, adultlike percentages of REM sleep from day one on (Hobson, 1989; Zepelin, 1994).

As a sidelight, this model of sleep as an erasing and rebooting process suggests a possible explanation for the anomalous size of dolphin brains. Dolphins don't appear to have REM sleep, and they exhibit slow-wave sleep on only one side of the brain at a time (Mukhametov, 1984). Being warm-blooded aquatic air-breathers, they can't afford to have breathing reflexes; they'd drown if they started breathing automatically while unconscious. So when they sleep, they hang motionless or swim slowly along at the surface and one hemisphere stays awake to breathe and locomote, while the other drops into slow-wave sleep. The other peculiar fact about dolphin brains is that they're amazingly big for animals that act so stupid. Small dolphins have brain and body weights resembling our own (Cartmill, 1990); but their behavioral repertoire and general intelligence seem somewhat subhuman—comparable to, say, those of a chimpanzee. Perhaps dolphins need those big brains because each half sometimes has to function entirely on its own while the other half sleeps. This may also explain why each cerebral hemisphere in a dolphin has its own totally independent blood supply, and why the commissural connections between the two hemispheres are relatively tiny (Ridgway, 1986).

I want to bring things back to the beginning now, and return to the title of this talk, by reexamining the question of animal consciousness in the light of what we know about the structured unconsciousness of sleep. There are three basic operating states of the healthy human brain: (1) waking consciousness, with alpha rhythms, mental events, and awareness of the world; (2) slow-wave sleep, with no mental events or awareness; and (3) REM sleep, with no awareness of the world, but with alpha rhythms and hallucinatory mental events. In REM sleep, the brain is partly disconnected from the body to inhibit responses to those hallucinations.

Although sleep has some secondary ecological functions, which vary from species to species, the main needs it serves appear to be those of the brain. Evidently, sleep restores something that is damaged or depleted by being conscious, or by things that we do when we are conscious. Animals that are (probably) never conscious don't sleep; animals that we know are sometimes conscious—that is, people—are compelled to sleep. So are the other animals that we believe for various reasons may be conscious (i.e., mammals and birds). Moreover, their sleep resembles ours in detail. The sleep of birds is different from ours in some features, as you might expect in a trait evolved in parallel; but in most mammals, states (2) and (3) are the same as ours in every respect.

It seems accordingly reasonable to think that state (1) in these animals is also the same as ours, that it includes mental events and awareness of the world, and that the subjective differences for them between these three states parallel our own as closely as the objective differences do. If restorative theories of the function of sleep are correct—if sleep is in effect the price we pay for freedom of the will—then animals that sleep as we do must also sometimes wake up as we do and experience their presence in the world.

Because we can't directly observe the contents of animal minds, the evidence for animal consciousness is necessarily indirect. But it seems at least as persuasive as the indirect evidence that we have for other unobservable phenomena—for example, the Big Bang, or neutrinos, or human evolution. The philosophers and scientists who refuse to acknowledge that dogs feel pain when you kick them seem to me to suffer from the same kind of ingeniously willful ignorance that we see in a creationist who rejects the notion of evolution because he has never seen a fish turn into a chicken. I am inclined to believe that these philosophers and scientists are not so much concerned about understanding the universe as they are about looking tough-minded and spurning the temptations of anthropomorphism.

To most of us, the temptations of anthropomorphism don't look quite so dangerous as all that. Our close animal relatives, after all, *are* anthropomorphic in the literal sense of the word, which means "human-shaped." They have organs like ours, placed in the same relative positions. And interestingly enough, they seem to recognize the same correspondences we do. Despite the conspicuous differences in sight, feel, and smell between a human body and a dog's, a friendly dog will greet you by licking your face and sniffing your crotch, and a murderously angry dog will go for your throat—just as they would behave in similar moods toward members of their own species. These are sophisticated homology judgments; and they encompass not only anatomy, but behavior as well. Just as we anthropomorphize dogs, horses, and other animals, they cynomorphize and hippomorphize us—and each other—right back in the other direction.

Psychological accounts of these facts often treat them as mistakes—category errors resulting from what Hediger (1950, 1981) has dubbed the "assimilation tendency" in social animals. I suggest that the assimilation tendency isn't a mistake, but an accurate perception of the way things are. In a world inhabited by closely related species, it confers an adaptive advantage. A gazelle that can tell when a lioness is thinking about hunting is less likely to be eaten; a lioness that can tell when a gazelle is thinking about bolting is less likely to go hungry. A man who doesn't notice that a horse is furiously angry, or a horse who can't make that sort of judgment about a human being, is correspondingly less likely to have offspring. Insofar as anthropomorphism recognizes and incorporates these facts about the world, it is not a vice but a survival skill. Indeed, one of the adaptive advantages of consciousness itself may lie precisely in the fact that it facilitates the reciprocal perception of other minds by analogy with our own—not just in our own species (Humphrey, 1987), but in others as well. If the construction of other minds in this way is both realistic and adaptively advantageous, as I believe that it is, then it is time to stop resisting its incorporation into the world view and vocabulary of science.

REFERENCES

Amlaner, C. M., Jr., and N. J. Ball

1989. Avian sleep. *In* M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and practice of sleep medicine, pp. 50–63. Philadelphia: Saunders.

Barlow, H. B.

1953. Summation and inhibition in the frog's retina. J. Physiol. 119: 69–88.

Barlow, H. B., R. Narasimhan, and A. Rosenfeld

1972. Visual pattern analysis in machines and animals. Science 177: 567–575.

Cartmill, M.

1990. Human uniqueness and theoretical content in paleoanthropology. Int. J. Primatol. 11: 173–192.

Crick, F., and G. Mitchison

1983. The function of dream sleep. Nature 304: 111–113.

Dennett, D. C.

1987a. The intentional stance. Cambridge: The MIT Press.

1987b. Consciousness. *In* R.L. Gregory (ed.), The Oxford companion to the mind, pp. 160–164. Oxford: Oxford Univ. Press.

1991. Consciousness explained. Boston: Little, Brown.

1995. Animal consciousness: what matters and why. Social Research 62: 691–710.

Gould, S. J.

1981. The mismeasure of man. New York: Norton.

Guilleminault, C., and W. C. Dement
1977. Amnesia and disorders of excessive daytime sleepiness. *In* R. R. Drucker-Colin and J. L. McGaugh (eds.), Neurobiology of sleep and memory, pp. 439–456. New York: Academic Press.

Hartmann, E.
1983. Two case reports: night terrors with sleepwalking—a potentially lethal disorder. J. Nerv. Ment. Dis. 157: 75–98.

Hartse, K. M.
1989. Sleep in insects and nonmammalian vertebrates. *In* M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and practice of sleep medicine, pp. 64–73. Philadelphia: Saunders.

Hediger, H.
1950. Wild animals in captivity: an outline of the biology of zoological gardens. London: Butterworth.
1981. The Clever Hans phenomenon from an animal psychologist's point of view. Ann. New York Acad. Sci. 364:1–17.

Heyes, C., and A. Dickinson
1993. The intentionality of animal action. *In* M. Davies and G. W. Humphreys (eds.), Consciousness, pp. 105–120. Oxford: Blackwell.

Hobson, J. A.
1989. Sleep. New York: Scientific American Library.

Humphrey, N. K.
1987. The uses of consciousness. 57th James Arthur Lecture on the Evolution of the Human Brain. New York: American Museum Natural History.

Jerison, H. J.
1973. Evolution of the brain and intelligence. New York: Academic Press.

Mahowald, M. W., and C. H. Schenck
1989. REM sleep behavior disorder. *In* M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and practice of sleep medicine, pp. 389–401. Philadelphia: Saunders.

Morgan, C. L.
1894. An introduction to comparative psychology. London: Walter Scott.

Mukhametov, L. M.
1984. Sleep in marine mammals. Exp. Brain Res. (Suppl.) 8: 227–238.

Rauch, P. K., and T. A. Stern
1986. Life-threatening injuries resulting from sleepwalking and night terrors. Psychosomatics 27: 62–64.

Regan, T.
1983. The case for animal rights. Berkeley: Univ. California Press.

Reite, M. L., K. E. Nagel, and J. R. Ruddy
1990. The evaluation and management of sleep disorders. Washington: American Psychiatric Press.

Ridgway, S. H.
1986. Physiological observations on dolphin brains. *In* R. J. Schusterman, J. A. Thomas, and F. G. Wood (eds.), Dolphin cognition and behavior: a comparative approach, pp. 31–60. Hillsdale, NJ: Lawrence Erlbaum Associates.

Searle, J. R.
1992. The rediscovery of the mind. Cambridge: The MIT Press.
Siegel, J. M., P. R. Manger, R. Nienhuis, H. M. Fahringer, and J. D. Pettigrew
1996. The echidna *Tachyglossus aculeatus* combines REM and non-REM aspects in a single sleep state: implications for the evolution of sleep. *J. Neurosci.* 16: 3500–3506.
Thorpy, M. J.
1990. Disorders of arousal. *In* M. J. Thorpy (ed.), Handbook of sleep disorders, pp. 531–549. New York: Marcel Dekker.
Turing, A. M.
1950. Computing machinery and intelligence. Mind 59: 433–460.
Webb, W. B.
1974. Sleep as an adaptive response. Percept. Motor Skills 38: 1023–1027.
Weiskrantz, L.
1986. Blindsight. Oxford: Oxford Univ. Press.
Wittgenstein, L.
1958. Preliminary studies for the "philosophical investigations," generally known as the blue and brown books. Oxford: Blackwell.
Zepelin, H.
1994. Mammalian sleep. *In* M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and practice of sleep medicine, 2nd ed., pp. 69–80. Philadelphia: Saunders.