THE NATURE OF MIND AND SELF

**"IN SEARCHING OUT THE TRUTH BE READY FOR THE UNEXPECTED,
FOR IT IS DIFFICULT TO FIND AND PUZZLING WHEN YOU FIND IT."**
— Heraclitus (C.535-470 BCE)

[24] THINKING THINGS

DUALISM AND PHYSICALISM

The 17th century French philosopher and mathematician René Descartes described the mind as a *res cogitans*, or "thinking thing." The mind is the thing that thinks, that also feels and desires — in a word, it's the thing that experiences. Experiencing is what minds do; or perhaps we should say: "that's how minds are." There is really no question that minds exist in some form or other — their existence is a commonplace of human experience. They are, it seems, where human experience quite literally takes place — unless the mind just is the collection of experiences, as opposed to a *thing* having an experience. The question here isn't whether minds exist but rather *what* they are.

Some people believe minds are the sort of thing that can exist wholly separate from a material body — we might want to call this kind of mind a "soul" and those who believe that minds are souls we can call "**dualists.**" René Descartes, for instance, was a dualist.

Other people believe minds are simply a certain way that certain kinds of bodies function or behave — that my mind is what I call my body, or a part of it, when it's having experiences or thinking or willing. This second group — we can call them "**physicalists**" — think that minds exist in much the same way that smiles exist. For instance, if you were creating an inventory of your face, you would list things like two eyes, two eyebrows, a nose, a chin, lips, perhaps a scar or two, and so on, but you probably wouldn't include 'smile' on your list — not because smiles don't exist or because you never smile, but because smiles don't exist in the same way that teeth and eyelids exist; they don't exist as some distinct part of the face. A smile is simply one way that a face can be organized or appear or behave. Except for the Cheshire cat in *Alice in Wonderland*, smiles don't exist apart from the face they are on. A smile is just a certain way that these various facial parts align themselves, or move together; it's more like a facial event than a facial part.

Physicalists maintain that minds are just like smiles. Of course minds exist, but not as something separate from the body. The mind is just a certain way that a body is organized or appears or behaves. If a person has a facial paralysis, he might not be able to smile. What he lacks is an ability, not a thing. Similarly, if a person is unconscious, what he lacks is an ability, not a thing; and a dead body is even more lacking in this regard. This is a physicalist understanding of the mind. On this view, the mind is just a certain way that a certain kind of body is able to function or behave.

Dualism and physicalism are not the only possible ways of thinking about the mind, but they are the most prominent and most basic, and so we will be focusing on them in this section.

ME AND MY MIND

A distinct but closely related question about the mind is its relationship to the self: How is my mind related to me, and how is your mind related to you?[1]

Is my mind just me? *Am* I a mind? Or do I *have* a mind? When I say: "Please hand me that pencil," I am presumably wanting the pencil given to my body, not to my mind as such — what would it do with a pencil, anyway? When considering these practical situations, the 'I' or the 'me' seem very much to be the mind/body composite, the organism as a whole.

Is my mind what perceives and thinks about the world? We certainly don't say things like: "my body saw a sparrow fly out of that bush" — but it sounds almost as strange to say that "my mind saw a sparrow…". It would be reasonable to interrupt anyone speaking like that to ask: "Do you mean that *you* saw the sparrow? What's all this talk of your *body* seeing or your *mind* seeing?"

It is clear, in a naïve sort of way at least, that one needs a mind in order to do things like think, wonder, believe, or doubt — and that one might get on well enough doing these things without a body — but that one definitely needs a body in order to do things like swim, play hopscotch, or digest one's lunch. Do these "normal ways of talking" tell us anything about what we *really* are?

CONSCIOUSNESS AND CAUSALITY

Try this experiment: imagine a 3-inch cube of wood painted on all sides with red paint. Now imagine the cube cut into 1-inch cubes. How many cubes will there be? And how many of these will have (i) three red sides, (ii) two red sides, (iii) one red side, (iv) no red side?

Most people tend to solve this problem by imagining the red cube being cut up, and then "visually inspecting" each of the smaller cubes in one's imagination. But what is it that solves the problem? The mental manipulation of these images, of which I am conscious? Or the brain processes, of which I am unconscious, that underlie these images?

Is *any* problem solved by way of our conscious thoughts and images? Or is all the work done by subconscious machinery in the brain underlying these thoughts and images? Does the physical event cause the non-causal mental event (a theory known as epiphenomenalism)? Or are the physical and mental "events" just two ways of describing the same event (an identity theory of the mind and brain)?

When we stop to consider the mind (is it me, or is it my mind, that does the considering?), we normally have in mind that part of us that is conscious or aware, the part that senses or perceives, and also that thinks — and that's why the following story is so peculiar.

A certain patient, known in the psychology literature as L.B., was having trouble seeing. It turns out that a tumor had destroyed part of his optical cortex, which is the part of the brain responsible for processing visual information. As a result of this damage, L.B. reported that he could see nothing on the left side of his visual field.

Nonetheless, when asked to guess where an object in his left field was, he would point correctly over 90% of the time. This suggested that there were neural pathways bypassing that part of the brain responsible for awareness, and yet which supplied perceptual information about the world. The visual data became part of the general background information available to the brain, even though the conscious subject was unaware of the data. This phenomenon is known as **blindsight**.

IMAGE ROTATION

Psychologists have found that people can rotate images anywhere from 320° to 840° per second, depending on the object rotated (for instance, letters and numbers can be rotated more quickly than other figures), as well as the age of the subject. Roger Shepherd, who worked with image rotation in the early 1970s, discovered a precise linear relationship between the angle an image is rotated and the time it takes to rotate it.

It has also been found that pigeons are able to solve these problems at the same speed, regardless of the degree of rotation (and they can do this more quickly than human beings).[2]

---

[1] An additional complication that we can't pursue here: What is it about my mind that makes it *mine* and not yours? And what is it about my thoughts that make them mine and not yours — for example, my thought that "5 x 7 equals 35", my desire to go back to bed, my memories of my 18th birthday? Can we share the same thought? If we are both drinking from the same bottle, are we tasting the same thing? If we are both contemplating the Pythagorean theorem, are we contemplating the same thing?

[2] See R. N. Shepherd and J. Metzler, "Mental Rotation of Three-Dimensional Objects" in *Science* 171 (1971) 701-3.

This story makes clear at least two things: First, that the status of the brain generally has a pronounced effect on the status of our experiences. This is something humans have understood for centuries, although we are only now developing some sense of the causal details involved. Second, it is possible to sense without being aware of the sensing. Is it perhaps also possible to think without being aware of the thoughts? If so, what role does consciousness play? Does it have a *causal* role?

When thinking about the mind we are immediately confronted with two contrasting points of view — the inner and the outer — both of which seem absolutely compelling, yet both of which, seemingly, cannot be correct. The mind would seem to inhabit this non-physical inner realm: My thoughts are in my mind, and they seem to be nowhere in space, suggesting that my mind is also nowhere in space. My thoughts would seem to lack all physical qualities, and thus my mind as well — and yet it is this very mind that allows me to perceive and to consider the physical world around me.

SUBSTANCE OR ATTRIBUTE?

Questions of free will and personal identity (and the possible survival of bodily death) depend on first deciding what the mind is. Does the mind exist as a distinct kind of substance? Or is it just an attribute of certain kinds of material bodies?

Zen Buddhism on the Self

"Why are you unhappy? Because 99.9 percent of everything you think and of everything you do is for yourself — and there isn't one."

— Wei Wu Wei, *Ask the Awakened* (1963)

We might ask what it means to "act freely" or to "be the same person over time," but ultimately these questions point to the more basic question of what the self or mind is. If physicalism is correct, and the mind is just a special way that the body functions — so, an attribute of the substantial body — then there is no *prima facie* reason for thinking that the mind might survive the death of the body. Similarly, there is good reason to believe that nothing can happen in the mind that is not causally related to earlier physical events in the body, thus making free will problematic.

[25] CARTESIAN DUALISM

René Descartes (1596-1650) developed a metaphysical view that involved two distinct kinds of substance: mental substances (the essence of which is to think), and material substances (the essence of which is to be extended). This view is what we call 'Cartesian Dualism.'


René Descartes
(France/Netherlands, 1596-1650)

According to Cartesian dualism, human beings are composites consisting of two distinct substances: the human body (a highly complex physical body) and a human mind (a simple soul). Cartesian dualism also claims that, in the context of the human body, mind and matter stand in causal interaction with each other. There is one human mind for each human body, and these two substances interact with each other. For instance, the mind experiences some sound coming from behind the body, and desires to have the body turn and look in that direction; here, the vibrations in the air strike the eardrums, causing a certain nervous excitation that travels to the auditory part of the brain, and ultimately "enters the mind" (or "I become aware of it"), at which point the sound occurs; the mind then directs the appropriate muscles of the body to contract or relax so as to turn the body in the proper direction.

This Cartesian world in which we live is actually two: a mental world in which minds exist with their ideas, and that is non-spatial and immaterial (and where each mind is connected to every other mind only indirectly, through their accompanying bodies), and a physical world in which bodies exist, extended in space, and where the material bodies are directly related to each other. My access to *my* mind is direct, but to *other* minds it is indirect.

THE APPARENT IRREDUCIBILITY OF THE MENTAL

Mental experience and mental terms do not seem to be reducible to the physical, and this irreducibility offers *prima facie* support for Cartesian Dualism. First, experience has a subjectivity or interiority to it that would seem to set it wholly apart from the physical world. We have *external sensations* (e.g., I see a red chair) and *internal sensations* (e.g., I feel pain), we have *mental imagery*, we suffer *emotions* (e.g., fear, anxiety, joy, sorrow, hope) — and all of this seems to occur inside us (not inside our bodies or brains, but rather inside the mind itself). For instance, when I eat a chocolate bar and experience the taste of chocolate, we assume that *something* is happening in my brain that makes possible that sensation of chocolate; but if a brain surgeon opened up my skull, there would be no part of my brain that she could lick and thereby have the same experience I am having. She might record neuron firings that *correspond* with my experience, but those firings seem to be quite different from the experience itself.

Along with this interiority of experience, three related and common beliefs and desires seem to recommend dualism. The first is the nearly universal belief that we are "free agents," that we are more than programmed robots or puppets on a string, that we can choose and deliberate and will our actions freely and decisively. Sometimes I choose to do something with my body *now* (this is actual willing); or I choose to do something on condition of some future event (this is conditional willing or intending). Yet if we are nothing more than bits of matter, then all of our thoughts and actions will be caused by the motions of other bits of matter, and our freedom will be wholly illusory. So human freedom, *prima facie*, seems to require metaphysical dualism.

A second feature is our feeling of personal continuity or identity. The matter of our bodies is always changing and, while our experiences are changing as well, there seems to be a continuity to our persons that transcends this change. Yet if we were only material beings, then such continuity and identity would seem to be compromised.

Related to this second feature is a third, the hope for immortality or an afterlife. If I am nothing but matter, then I will cease to exist once my material being disintegrates (such as when my body dies). If, on the other hand, I am an incorporeal, indivisible mental substance, then the death of my body is nothing to me, for the real self cannot die (the only way it could die is through disintegration; but if it is simple and indivisible, then it obviously can't be divided into parts, and so it cannot disintegrate). Admittedly, it is a standard part of most Christian confessions that one's body will be resurrected at some future time, thus allowing for one's continued existence. But that sort of immortality depends upon divine intervention, and so lacks the certainty and universal appeal of a proof that the self is an immaterial soul. (For more discussion of these issues of free will, personal identity, and the survival of death, see the chapters on "Free Will and Determinism" and "Personal Identity and The Afterlife," both below.)

DESCARTES' ARGUMENTS FOR DUALISM

Descartes offered several arguments for viewing mind and body as distinct substances. One was a result of his methodological doubt: I can imagine not having a body, but I cannot imagine not having a mind. Therefore mind must be separate from body, and while it may be true that I *have* a body, it is the case that I *am* a mind.[3]

A slightly better argument for dualism is to note that a material body is divisible, but mind would seem to be indivisible. That is, I can imagine taking a bit of matter (some body) and dividing it into pieces or parts; but I cannot imagine doing the same to a mind (or *my* mind). Minds have a unity about them not found in matter. Since everything that is extended is divisible, mind must not be extended; and if it is not extended, then it obviously is distinct from matter; thus it is a different substance.

This argument from the indivisibility of mind has two different forms. The first is *conceptual*: I cannot conceive of mind having any parts into which it can be divided.[4] Mind must be unified, for otherwise it could not have a

---

[3] This is a horrible argument. It fails to notice that we might know the same thing in more than one way, and thus entertain contradictory beliefs about it; for instance, humans used to believe that the morning star and the evening star were separate planets, when in fact they are both Venus, but appearing on different sides of the sun. Another example: if you didn't know that Mark Twain was a penname for Samuel Clemens, you could well hold the beliefs that Mark Twain was the greatest author who ever lived and that Samuel Clemens was not an author at all, much less the greatest.

thought. For instance, if one part of the mind began a thought, and another part of the mind completed the thought, then there would be no thought at all. It would be like having separate individuals each thinking one word of the proposition: here, the whole thought (e.g., "There's a red balloon in that tree") would not occur at all.

The second argument is *experimental*: although the mind seems to inhabit the whole body, we do not sever or divide the mind when we sever or divide the body, such as when a foot is amputated: this does not result in a corresponding amputation of the mind.

Problems with Dualism

Despite these various considerations in favor of a dualist understanding of the mind, philosophers have been quick to point out several problems with Cartesian dualism that appear to be very nearly intractable. I consider them separately, below, but they all center on the basic puzzle of how immaterial minds and material bodies are supposed to causally affect one another.

The conservation of matter and energy

It has been argued that any interaction between mind and body will violate the physical principle of conservation, for it opens up what was a closed physical system. On Descartes' account, minds appear to be adding energy to the material system whenever the mind moves the body to do something, and energy appears to be lost to the mind whenever the body affects the mind.

A Cartesian might reply that the principle does not apply to brain phenomena, or that there may not be any net gains or losses (it may take no energy for the body to act upon the mental, and the mental may be able to effect changes in matter that doesn't involve any addition in energy).

Leibniz on Cartesian Interactionism

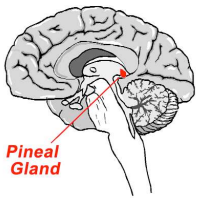
"Descartes recognized that souls cannot impart any force to bodies, because there is always the same quantity of force in matter. Nevertheless he was of the opinion that the soul could change the direction of bodies. But that is because in his time it was not known that there is a law of nature which affirms also the conservation of the same total direction in matter. Had Descartes noticed this he would have come upon my system of pre-established harmony."

— G. W. Leibniz, *Monadology*, §80

How can minds and bodies interact causally?

Mental and material substance are so dissimilar that it is wholly unclear how they are supposed to causally interact with one another. We understand how two bodies interact: one bumps into the other, and causes it to move. This mechanical interaction is the sort of account that Descartes tried to give of the workings of our bodies. But the body cannot "bump" into the mind because there is nothing physical that it can bump into. Minds will offer no resistance to the bodies; similarly, the mind cannot "bump" into a body.

In short, the causal interaction between my mind and my body — which, according to Descartes, is supposed to occur in the pineal gland — is wholly mysterious, and it is a mystery of the worst sort: not only do we not know how the interaction occurs, it appears that we can *never* know — it is, in principle, beyond our ken.



The apparent dependence of the mind on the brain

Whatever mind is, it seems to be closely dependent upon the condition and fate of the brain, which suggests that the mind is not a free-floating immaterial substance. When chemicals like alcohol are ingested, the mind is clearly affected — not just what it perceives, but how it operates and thinks. If the mind were a separate immaterial substance, one would think that the mind's operations would be safe from any changes to the brain, and the effects would be limited to whatever control it might have over the body or the ability of the brain to transmit information to the mind. Sensory information would be channeled through the brain, but since thinking is what the mind itself does, the thinking should not be impaired by the ingestion of alcohol or other "brain-altering" drugs. Similarly, too

---

4 One might, indeed, argue that the mind *does* have parts — after all, there are distinct abilities of thinking, feeling, and willing. But Descartes claims that each of these is performed by the whole mind.

little oxygen to the brain can cause a person to faint or "black out"; if dualism is true, one might imagine the sensory inputs being disturbed at this point, but it isn't clear why the mind would lose its ability to function at all — to remain conscious, to think, and so on — and yet this is what happens. A blow to the head disturbs the brain and its functioning, but why would it also disturb the mind and its functioning, if the mind truly is an immaterial substance? In short, the fates of my mind and my brain are so closely intertwined, they seem to be identical, or nearly so.

How are minds and bodies connected?

Closely related to the problem of causal interaction is understanding how individual minds and bodies are connected together. What is it that connects my mind to my body, and not to someone else's body? If mind is immaterial and non-spatial, it would seem as though it might end up connected to anything. What ties it down to *this* particular lump of matter?

Initially, one might suppose that there is some sort of physical connection. But this can't be right, since the mind is (by definition) non-physical. There isn't any obvious way it might get hooked to a physical thing, such as a neuron, or something like the pineal gland. Lacking a straight-forward physical connection, we might turn to a connection by virtue of occupying the same space or contiguous spaces. But this won't work, either, for while bodies are in space, and therefore have a location, minds are non-spatial.

In order to talk about the location of minds and mental events, one might develop a distinction between *local* and *virtual* placement in space: the mind is in the body *virtually* but not *locally*, that is, the mind seems to have a location, but not a precise one — for instance, I'm certain that my mind isn't somewhere on the moon. In fact, I'm pretty sure that my mind is somehow inside my body, and perhaps even inside my skull. But I'm not sure where, exactly, it is in the skull — maybe it is co-extensive with the brain. But we don't want to say that the mind is extended, for it seems to have a unity that resists extension. This distinction between *virtual* and *local* placement in space, however, really seems to be just a fancy way of saying that we traditionally attach our minds to our bodies, although we aren't sure how this is done.

[News clipping]

STUDY SAYS MALE BRAINS BIGGER THAN FEMALE BRAINS

COPENHAGEN, Denmark (AP) – Danish researchers say they've found that men, on average, have about 4 billion more brain cells than women. But they haven't figured out what men do with them.

Dr. Bente Pakkenberg, a Copenhagen Municipal Hospital neurologist who led the research project, told Danish radio last month that the conclusions came from an examination of the brains in 94 cadavers of people age 20 to 90.

The average number of brain cells in males was 23 billion, while the females had about 19 billion. Asked what the males might be doing with the surplus, Dr. Pakkenberg said: "Right now it's a mystery. The knowledge we already have shows men are not smarter than women."

American Medical News (August 18, 1997)

[26] PHYSICALISM

Dualism is the view that reality consists of two separate kinds of things: material bodies and immaterial minds, each with their corresponding events. Monism, on the other hand, claims that reality consists of one kind of thing, which is either mental or material. The only traditional view of idealistic monism is George Berkeley's (discussed in some detail in a previous chapter).

If we reject the Cartesian hypothesis that minds are "mental substances" separate from "bodily substances," then we could say that a mind is extended equally with its body, and that it is simply the way that the body functions (insofar as it thinks, feels, and desires). Here the unity of the mind is a "functional" unity (just like the unity found in a properly functioning automobile). This non-Cartesian approach, of course, rejects the notion that minds and bodies are separate (or even separable), and thus does not solve the problem of connecting minds and bodies so much as dissolves it.

We will now examine below various materialistic forms of monism.

SUPPORT OF MATERIALISTIC MONISM

Apart from the problems noted above that plague dualism, materialistic monism is further supported by the following considerations.

First, Cartesian dualism assumes a clean break between those mechanical bodies that have minds, and those that don't. Such a clean division, however, is belied by animal behavior, which indicates great similarities up and down the ladder of complexity, from human beings and other primates down to rats, birds, lizards, and worms. This was a problem pointed out even in Descartes' day: if non-human animal behavior is explicable in mechanical terms, then human behavior is as well, and vice versa. This continuum makes dualism highly suspect.

Second, Cartesian dualism results in a skepticism of other minds. (This is a problem for all dualistic theories.) As Gilbert Ryle muses, if Cartesian dualism is true, then "for all that we can tell, the inner lives of persons who are classed as idiots or lunatics are as rational as those of anyone else. Perhaps only their overt behavior is disappointing; that is to say, perhaps 'idiots' are not really idiotic…" (*The Concept of Mind*).

Finally, by segregating the mental world off as a separate substance, then psychology as a science becomes impossible. We cannot study other minds, since we cannot properly get at them (they are invisible, non-material, private, etc.). In the following, we will briefly consider three physicalist theories of mind.

[Poem]

#632

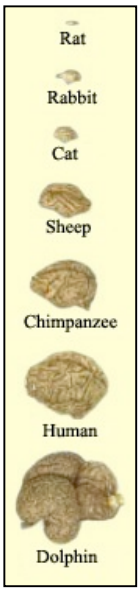
The Brain — is wider than the Sky —
For — put them side by side —
The one the other will contain
With ease — and You — beside —

The Brain is deeper than the sea —
For — hold them — Blue to Blue —
The one the other will absorb —
As Sponges — Buckets — do —

The Brain is just the weight of God —
For — Heft them — Pound for Pound —
And they will differ — if they do —
As Syllable from Sound —

— Emily Dickinson (1830-86)

MIND/BRAIN IDENTITY THEORY



Mind/Brain identity theory is the view that the mind just is the brain, or at least some part of it, and therefore that mental events are identical with certain physical events located in the brain. When a certain group of neurons fire in a certain way, that just is a visual image of a certain shade of red, or a certain feeling of sadness, or a memory of one's 12th birthday. Many physical events in the world have simply an outer or external aspect, but some events (many that occur within a brain) have an inner aspect as well as an outer aspect.

Identity theory, like Cartesian dualism, allows for us to think of the mind as a substance or thing. Unlike with dualism, however, the mind is now just a special kind of physical thing.

THE BRAIN

The adult human brain weighs, on average 1350 grams (about three pounds), and is about the size of your two fists pressed together. Our closest living cousins — the chimpanzees — have brains only one-third as large, while blue whales have brains five-times larger than ours. More significant, however, is not the absolute weight of a brain, but the brain/body weight ratio; here we find that the human brain is six-times as large as what would be expected from the ratio found in other mammals.

Brains are biologically expensive: with only 3% of the body's weight, they consume 17% of the total calories — this caloric requirement suggests a close relationship between the growth in human brain size and our diet during the course of our evolution.

Neurons — the cells that comprise about one-half the bulk of the brain — come in over 200 varieties, and there are about 100 billion of these cells in the brain. The interconnections among these neurons are estimated at 1000 trillion.[5]

---

[5] The image comparing the brains of different species, and much of the information in this box, comes from Bruno Dubuc's excellent website: "The Brain from Top to Bottom," sponsored by the *Canadian Institutes of Health Research: Canadian Institute of Neurosciences, Mental Health and Addiction* (http://thebrain.mcgill.ca), accessed June 21, 2011. See also Douglas Fox, "The Limits of Intelligence" in *Scientific American* (July 2011), 37-43.

Possible problems with identity theory involve the location of mental events and the apparent privileged access one has to one's own mental events. First, the mind and its thoughts don't seem to be located in space, whereas physical events are very much located in space, and if mental events are identical with certain brain events, then the mental events do indeed occur in space. This may not be much of a problem, however, since it trades on perhaps dubious intuitions, and in any event it would also seem that thoughts clearly do occur in space, since they seem to be taking place in one's head.

A second possible problem is that I seem to have a "privileged access" to my own mental events, whereas the physical events of my brain are essentially open for anyone suitably situated to observe. The identity theorist will claim that this seeming privacy of the mental is an illusion. The neurologist can see the process occurring that just is the event of thinking (believing, experiencing, etc.) something.

FUNCTIONALISM

Is the actual stuff making up the brain important for there to be a mind? The identity theorist thinks it *does* matter, since the mind just is the brain: If there is no brain, then there is no mind. The functionalist, however, disagrees. Imagine replacing the brain — neuron by neuron — with electrical linkages. A neuron collects electrical charges from other neurons, and passes these charges down the line to the next neuron. Without too much difficulty we might replicate this causal chain by using electrical wires and switches.[6] Functionalism is the view that such a project — at least in principle — could be successful. The physical material that "embodies" the mind is not important. What matters is the "causal array" of that embodiment, or its functional state.[7]

Functionalism is in some ways a cleaned-up version of behaviorism. It holds that we can define mental states in terms of their cause, the effects they have on other mental states, and the effects they have on behavior. The net result is that talk about mental states is ultimately reducible to talk about sensory inputs and behavioral outputs.

Mental events and physical events are different ways of describing the same system. Mental events are individuated by their causal or functional role within the brain. The mind is a causal array or network, and as such could be implemented in all sorts of materials, including brains.

Functionalism is a materialist theory of the mind that avoids the problems of correspondence that trouble the mind-brain identity theory. Functionalism involves distinguishing between *physical descriptions* and *abstract (functional) descriptions* of systems, that is, the rules governing a function, and the physical manifestation of those rules or function. The physical manifestation might occur in a brain or in a computer.

Similarly, we can describe the brain in two ways: *physically* (giving a description of the neurons and their interconnections and order of firings) and *functionally* (using mental terms primarily for describing the function of those certain operations). A certain event in the brain will be an act of thinking not because it is a special *kind* of brain event, but because it performs the appropriate *function* in the brain's program. Functionalism is closely related to work on artificial intelligence, to which we turn in the next section.

---

6 Researchers at the University of Lille (France) have recently accomplished something like this, developing organic transistors that mimic the synapse; see Dominique Vuillaume, *et al*., "An Organic Nanoparticle Transistor Behaving as a Biological Spiking Synapse" in *Advanced Functional Materials* 20 (2010): 330-37.

7 Admittedly, this mechanical mind (as described) would be static. To have new experiences, neurons need to keep forming new synapses, and re-enforcing or degrading old ones. So for this thought experiment to work, we need the mechanical replacements to be capable of re-aligning themselves — something more easily done at the software level than the hardware level, but certainly *possible* at the hardware level.

[27] ARTIFICIAL INTELLIGENCE: CAN COMPUTERS THINK?

ANIMAL BEHAVIOR, RATIONAL SOULS, AND CLEVER ROBOTS

I see these human beings walking about, interacting with each other and with myself: How do I know that they aren't just cleverly-built robots? Is there a test that would allow us always to know when we are confronted with a real "person" — a Cartesian thinking thing — instead of some programmed machine?

Descartes' metaphysical dualism implies that the human body, being made up entirely of matter, is just a complicated machine — divinely crafted, of course, but nonetheless a machine following mechanical laws. The human mind or soul inhabits this machine, and stands (in some mysterious way) in interaction with it, such that the mind "controls" at least some of what the machine does. Similarly, things that happen within or to the machine are often consciously experienced by the mind.

Descartes also believed that non-human animals ("brutes") were simply machines, and nothing more. He believed this on the basis of **two tests** that he describes in his *Discourse on Method* (1637). The ability to speak was Descartes' first test. He claimed that the absence of brute speech is not due to lack of speech organs (after all, magpies and parrots can imitate the human voice) — and even if they did lack these organs, we find that deaf and dumb human beings still create a language, unlike brutes. Further, human speech is more than mere "expression of passion," which is all that brutes are capable of performing. We must not suppose that brutes possess some "unknown language," Descartes argues, for if this were so, then they could communicate their thoughts to us as easily as they can to each other, and they clearly do not communicate their thoughts to us.

Descartes' second test is actually best viewed as his principle criterion, with speech being just an example. This test concerns the universality or adaptability in one's behavior. "Reason is a universal instrument," and thus can adapt to any contingency — for instance, developing novel strings of words for novel situations. Descartes found that various animals were exceptionally skilled at a few things — even out-performing human beings, just as an adding machine can add sums more quickly than we can. But while quite good at one or two skills, they perform horribly overall, since they are unable to adapt to the peculiarities of each new situation. (This is all quite false, of course, as the animal studies of the past century have shown; but such were Descartes' beliefs.)

The implications of Descartes' arguments are fairly severe. If non-human animals fail these tests, then they are understood to lack souls; and if they lack souls, then they lack mental lives, and so are fundamentally no different than human built machines, like clocks or calculators. They cannot think, nor can they suffer.

At least two questions confront us here: (1) Are these tests a proper indication of the presence of a rational mind? and (2) Can non-human animals truly not pass them? These tests were questioned from the very start, and some of Descartes' contemporaries turned his argument in the opposite direction: Because animal behavior did not seem all that different from what humans do, if all animal behavior could be understood mechanistically, then so could all human behavior — and thus we should think of ourselves as nothing more than machines. The most famous proponent of this view was the French philosopher and physician **Julien Offray de La Mettrie** (1709-1751) and his notorious book, *Man a Machine*.[8] Drawing a clear line between human beings and other animals has not been easy, and it is constantly being redrawn as we increase our understanding of other animals. We once thought that only humans could use tools, or could pass down information from one generation to the next, or engage in play, or deceive others, or form concepts, or have a "theory of mind" (a sense of the intentions of another individual). Each of these lines was eventually erased by ethologists and comparative psychologists who study the behavior of other animals.

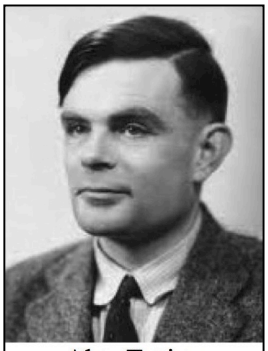

Julien Offray de La Mettrie
(France, 1709-1751)

As it turns out, there actually are two lines to draw, not one — although this has not always been clear in the history of the discussion. First, we are looking for an essential difference between human beings and other animals; second, we are looking for an essential difference between human beings and humanly-built computers and robots.

---

8 Julien Offray de LaMettrie, *L'homme machine* (Leyden, 1748).

These are potentially quite different borders to negotiate, and I would like now to turn exclusively to a consideration of the latter border.


Alan Turing
(Great Britain, 1912-1954)

Alan Mathison Turing (1912-1954) was an English mathematician, logician, and early theorist of computer science who, among other things, built a computer used to crack the German military code (devised by their own "Enigma" machine) during World War II.

Turing was also interested in the field that is now called "artificial intelligence," and he developed the famous **Turing Test** as a criterion for deciding whether computers can indeed think.[9] This test was actually quite simple: it involved two humans, A and B, and a computer, C. The first human, A, would communicate, by way of a keyboard, with B and C. A would ask any question he liked of his two interlocuters, and if he was unable to reliably say which was the human and which the computer, then the computer was said to have "passed the test" and, for all practical purposes, would be said to be in possession of a mind (i.e., be able to *think*). It is with the articulation of this test that the field of artificial intelligence officially began.

TURING MACHINES

If it's in state:	1			2			3			4		
and it reads:	A	-	B	A	-	B	A	-	B	A	-	B
then it writes:	A	-	B	B	-	B	A	-	B	A	-	A
and moves:	r	-	r	-	-	r	r	r	l	r	-	-
and enters state:	1	1	2	3	2	2	4	4	3	1	4	4

Turing machines are the basis of all computers that exist today. The hardware to be used is left unspecified; a Turing machine could be implemented in a structure made of banana peels and egg shells, although perhaps with some difficulty. Normally, silicon chips are used to implement them. They are characterized as having a finite number of states, where a **state** is a disposition to act. The possible actions are to read a symbol, erase and/or write a symbol, move to an adjacent cell (either left or right) to read another symbol, and change to a different state. The **symbols** could be thought of as existing on a long tape, but they could just as easily be embodied in a number of different media, such as iron oxide dust on a floppy computer disk or pits in the surface of a DVD. Depending on the state that the machine is in and the symbol that is being read, the machine will perform any of the following **actions**: (i) move to the previous or next symbol, or continue reading the same symbol; (ii) erase the symbol and write another symbol; and (iii) change to a different state, or remain in the same state. The sample machine in the accompanying box is designed to take any string of A's and B's (our sample symbols) and re-order them so that all the A's come first, followed by all the B's. It's a simple machine (much simpler than one designed to add or subtract numbers), but it does its job transparently and well. It consists of four different states, which are described in terms of how the machine responds when it reads a certain symbol (A, B, or no symbol). Imagine a sample tape with the letters 'BABA', and now imagine moving between the four states of the machine, as described in this table, as you grind through the letters of the sample tape (begin in state 1 reading the 'B' on the far left). After fifteen or so moves, the sequence 'BABA' will be re-ordered as 'AABB' and the machine will stop.

MACHINE STATES AND STATES OF MIND

The view that the mind is just a fancy Turing machine is rather compelling. The states of Turing machines can be thought of as "dispositions to behave" just as minds have dispositions. If a Turing machine is in state #1, for instance, and it sees a "0", then it might erase the "0" and write a "1", move to the next symbol, and enter state #2. If I am in a hungry state and I see a pizza, then I might move to the pizza, consume a portion of it, and enter the state of satiation.

---

9 Alan Turing, "Computing Machinery and Intelligence" in *Mind* 59 (1950): 433-60.

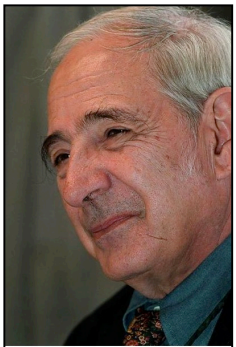
Artificial intelligence (AI) is the attempt to simulate human intelligence in a computer. It assumes a functionalist account of the mind — the mind is just the functional description of the body, primarily the brain. Therefore this function might, in theory, be replicated or modeled in a computer (thus producing artificial intelligence).

If a task can be done on a Turing machine, then that task is **algorithmic** (or computable). This is "Turing's Thesis," and was the first precise definition of what an algorithm is. A task is algorithmic, in other words, if the process for performing the task is so well defined that a mere machine can do it. It is hard to know whether a task is algorithmic until you attempt to program it onto a computer. For our purposes, the question is whether everything that the mind does is also algorithmic; if it is, then we should be able to implement or model the mind in a computer. At that point, it *might* be legitimate to say that the computer can think.

Artificial Intelligence as a "Top-Down" Strategy

One can try to explain what the mind is in either of two general ways: from the bottom-up or from the top-down. Bottom-up strategies begin with the "atoms" of mental experience and work upwards until reaching the complex phenomena of various mental skills (such as remembering, learning, and pattern-recognition). The two likeliest candidates of this bottom-up strategy are behaviorism (focusing on stimuli and responses) and a neuro-physiological approach that looks at firing patterns of individual neurons. Each of these comes with its problems: the stimuli and responses that behaviorism acknowledges aren't likely to be the relevant atoms, and with neurophysiology, there are so many neural connections that, even while these are likely our best candidate for the "mental atoms," the technical difficulties surrounding their exhaustive study appear to be, at least at present, insurmountable. These problems make top-down strategies more attractive. With this top-down approach, you analyze complex mental phenomena into ever smaller units of organization until you arrive at non-conscious elements (such as neurons and their connections). This strategy best characterizes AI and traditional epistemology — for instance, the most general top-down approach is Kant's: How could anything experience or know anything?

One general strategy in AI is to analyze our mental functions into simpler and simpler functions until finally the functions, when viewed by themselves, no longer appear to be minded or intelligent. Consider the problem of how we form a visual representation of the world. A naïve view of this process, put as crudely as possible, assumes that there is a person inside your brain that interprets the images coming in, as though there were a movie screen inside the head (these are the internal representations), as well as a little person (or *homunculus*) watching the show (that is, interpreting these representations). This account, however, does little to explain how we understand the world; it just puts the problem off a step, for either the homunculus understands what he sees or he does not; if he does not, then neither do we; if he does, then there must be an even smaller homunculus inside of him, observing its own set of internal representations (and here, of course, we enter an infinite regress). Representations cannot simply understand themselves; there must be an interpreter. The approach of AI is to solve this problem by breaking down this interpreter-function into sets or structures of functions that are so simple that they do, in fact, understand themselves. The mind, as we know it, disappears into its non-mental parts, becoming nothing more than the sum-total of these parts insofar as they are functioning together.[10]


John Searle
(USA, 1932-)

Searle's Criticisms of Artificial Intelligence

John Searle (b. 1932) teaches philosophy at the University of California/Berkeley and has become a prominent critic of functionalism and the AI project. In his essay, "The Myth of the Computer" (1982), Searle notes that there are three levels for explaining human behavior. The first level is what has come to be called "**Folk psychology**," the common-sense understanding of conscious intelligence. This consists of hundreds of common-sense generalizations or laws like "Persons in pain tend to want to relieve that pain" or "Persons who are angry tend to be impatient." These laws make use of various concepts like belief, desire,

[10] Cf. William Lycan's "homuncular functionalism" as discussed in his "Form, Function, and Feel," *Journal of Philosophy*, 78 (1981) 24-49.

fear, and pain, and we use these laws and concepts to explain and predict human behavior. This level of explanation works well enough in practice, but is not scientific.

In the past several centuries, Searle notes, we have become convinced that our folk psychology is somehow grounded in the workings of the brain. Neurophysiology — a second level for explaining human behavior — is scientific, but not well developed, and (perhaps merely as a consequence of its immature state) it cannot explain much of our behavior.

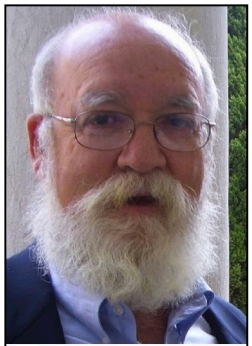
Cognitive science is the most recent attempt at a third level between these two — a kind of a scientific psychology that is not introspective, and yet not merely a study of the brain.

Many cognitive scientists see at the heart of their field a theory of mind based on artificial intelligence, that Searle summarizes with three propositions: (1) the mind is a program, (2) the neurophysiology of the brain is irrelevant, and (3) the Turing test is the criterion of the mental. Searle criticizes each of these propositions. Against the claim that **the mind is a program**, Searle notes that the mind does one thing that no program does: it attaches an interpretation to the symbols used. As Searle puts it, computer programs are mere **syntax without semantics**; the symbols remain uninterpreted in the computer. Searle supports his criticism with what has become a famous thought-experiment: **the Chinese Room**. He asks us to imagine a room without windows, but with something like two mail slots — one for incoming pieces of paper, and one for outgoing — and hundreds of books lining the walls inside the room. The room also contains one non-Chinese speaking human adult — call her Betty. The pieces of paper sent into the room contain sentences written in Chinese, and the books are filled with transformation rules that tell Betty how to respond (also in Chinese) to these sentences. Betty need not know that the sentences are in Chinese, or even that they are sentences. All she needs to do is identify the string of symbols in one of the books and then copy out the corresponding set of symbols that the book indicates. Now suppose that a Chinese speaker, Wenje, is writing down messages and sending them into the room, and that appropriate responses are coming back out. It would appear that Wenje is having a conversation with Betty. But by hypothesis, Betty doesn't know that the symbols she is manipulating are sentences, much less Chinese sentences, and she has no idea that she is conversing with someone. But this is precisely the situation of a computer: It shuffles symbols around following pre-set rules (the syntax), with no understanding (the interpretation or semantic content of the symbols) of the symbols. Therefore, the computer has no semantics, no understanding of the symbols.

The second proposition — that **the neurophysiology of the brain is irrelevant** — seems to rest on the notion that a computer simulation is the same thing as whatever is being simulated. If we can manage to simulate the workings of the brain on a computer, then there is nothing significantly different between the two. But Searle finds this absurd. A computer might simulate the various mechanisms involved in our feeling thirsty, and even have it print out the words: "I'm thirsty" — but no one would contend that the computer really is thirsty. Much of our behavior, Searle continues, is grounded in the kind of physical beings that we are, not simply in the way that these beings function.

Searle is being tendentious here. His examples seem crazy, because computers aren't the sort of things that eat or drink (and thus are not the sort of things that get thirsty or hungry). But strong AI doesn't claim that computers are beings capable of thirst or hunger; rather, it claims that they are capable of *thought*. Thirst needs a body, but does *thinking* need a brain? Strong AI does not think so; but Searle disagrees:

> I believe that everything we have learned about human and animal biology suggests that what we call "mental" phenomena are as much a part of our biological natural history as any other biological phenomena... Much of the implausibility of the strong AI thesis derives from its resolute opposition to biology.


Daniel Dennett
(USA, 1942-)

Finally, Searle believes that his Chinese room thought-experiment undermines **the Turing test**. Wenje, the native Chinese speaker, might easily believe that he is having a conversation with someone who understands Chinese, when by definition he is not.

Searle's arguments against AI have not gone unchallenged. **Daniel Dennett** (b. 1942) and others have argued that the Chinese Room argument fails to undermine AI because it mistakes the level at which "understanding" takes place. In the Chinese Room, Betty clearly has no understanding of Chi-

nese, or even what she is doing — that's true by the very terms of the argument. But Dennett wishes to argue that the room itself understands Chinese. This is the "systems reply" to Searle — a reply that Searle finds preposterous. When put in terms of the thought-experiment, the systems reply might indeed seem preposterous, but Dennett would argue that this preposterousness is only an illusion caused by the terms of the argument. After all, we have entities who are clearly conscious beings — Betty, Wenje — and it's also clear that Betty understands none of the Chinese being spoken, whereas Wenje does. Because they are both (*ex hypothesi*) human beings, then it would seem that they are at the same epistemic level — but of course they are not. The entire Chinese Room is at the same level as Wenje, and inside Wenje we could postulate some analogous Betty who is equally oblivious to what is going on.

What do you think?

READINGS

DISCOURSE ON METHOD (SELECTION)

René Descartes

It is difficult to exaggerate the importance of René Descartes (1596-1650) to the history of modern science and philosophy. It was Descartes, for instance, who developed analytic geometry, the mathematical key to the development of modern physics. One of the earliest of Descartes' publications, written in French, was his Discourse on Method for Conducting One's Reason Rightly and for Searching for Truth in the Sciences, *published in 1637. This short work was divided into six parts, and served as a methodological preface for three treatises on optics, geometry, and meteorology. Part Five of the* Discourse *summarizes a longer work of his,* Le Monde *(The World), that he was about to publish five years earlier, but then suppressed after news reached him of Galileo's trial in Rome. Here Descartes develops a mechanistic view of nature, including the claim that all animals (other than human beings) are nothing more than divinely crafted machines. In the following brief selection from Part Five, Descartes gives an account of the two tests that determine whether or not a being has a rational soul.*

If there were such machines having the organs and the shape of a monkey or of some other nonrational animal, we would have no way of telling whether or not they were of the same nature as these animals; if instead they resembled our bodies and imitated so many of our actions as far as this is morally possible, there would still remain two most certain tests whereby to know that they were not therefore really men. Of these the first is that they could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others: for we may easily conceive a machine to be so constructed that it emits vocables, and even that it emits some correspondent to the action upon it of external objects which cause a change in its organs; for example, if touched in a particular place it may demand what we wish to say to it; if in another it may cry out that it is hurt, and such like; but not that it should arrange them variously so as appositely to reply to what is said in its presence, as men of the lowest grade of intellect can do.

The second test is, that although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it could be discovered that they did not act from knowledge, but solely from the disposition of their organs: for while reason is an universal instrument that is alike available on every occasion, these organs, on the contrary, need a particular arrangement for each particular action; whence it must be morally impossible that there should exist in any machine a diversity of organs sufficient to enable it to act in all the occurrences of life, in the way in which our reason enables us to act.