This was a review of John R. Searle’s The Rediscovery of the Mind (M.I.T. Press, 1992). Though Searle and I agree about a great deal, I don’t think it’s possible to distinguish his antireductionist solution from property dualism. And I do not believe it could be a brute fact of nature that the higher order mental properties of the nervous system should be produced by the details of its physicochemical operation. The relation between the levels must be more “internal” than that—a form of intelligibly necessary consequence rather than pure correlation. The irreducibility of the ontologically subjective to the ontologically objective continues to be an obstacle to the imaginability of such a connection.

According to a widely held view, the brain is a giant computer and the relation of the human mind to the human brain is like that of a computer program to the electronic hardware on which it runs. The philosopher John Searle, a dragonslayer by temperament, has set out to show that this claim, together with the materialist tradition underlying it, is nonsense, for reasons some of which are obvious and some more subtle. Elaborating arguments that he and others have made over the past twenty years, he attacks most of the cognitive science establishment and then offers a theory of his own about the nature of mind and its relation to the physical world. If this pungent book is right, the computer model of the mind is not just doubtful or imperfect, but totally and glaringly absurd.

Searle’s book begins with a lucid critical survey of the different views now circulating about the relation of the mind to the body. The mind-body problem was posed in its modern form only in the seventeenth century, with the emergence of the scientific conception of the physical world on which we are now all brought up. According to that conception, the physical world is in itself colorless, odorless, and tasteless, and can be described mathematically by laws governing the behavior of particles and fields of force in space and time. Certain physical phenomena cause us to have perceptual experience—we see color and hear sound—but the qualities we experience do not belong to the light and sound waves described by physics. We get at the physical reality by “peeling off” the subjective effects on our senses and the way things appear from a human point of view, consigning those to the mind, and trying to construct an objective theory of the world outside our minds that will systematically explain the experimental observations and measurements on which all scrupulous observers agree. However radically the content of contemporary physics and its conception of the role of the observer may differ from that of classical physics, it is still in search of a theory of the external world in this sense.

But having produced such a conception by removing the appearances from the physical world and lodging them in the mind, science is faced with the problem of how to complete the picture by finding a place in the world for our minds themselves, with their perceptual experiences, thoughts, desires, scientific theory construction, and much else that is not described by physics. The reason this is called the mind-body problem is that what goes on in our minds evidently depends on what happens to and in our bodies, especially our brains, yet our bodies are part of the “exter-
From the properties that are being reduced. When we say heat consists of molecular motion, we mean that heat as an intrinsic property of hot objects is nothing but the motion of their molecules. Such objects produce the feeling of heat in us when we touch them, but we have expressly not identified that feeling with molecular motion—indeed, the reduction depends on our having left it out.

Now how could mental phenomena be reduced to something described entirely in physical, nonmental terms? In this case, obviously, we cannot leave out all effects on the mind, since that is precisely what is to be reduced. What is needed to complete the materialist world picture is some scheme of the form, “Mental phenomena—thoughts, feelings, sensations, desires, perceptions, and the like—are nothing but . . . .” where the blank is to be filled in by a description that is either explicitly physical or uses only terms that can apply to what is entirely physical. The various attempts to carry out this apparently impossible task and the arguments to show that they have failed make up the history of the philosophy of mind during the past fifty years.

Scarle’s account of that history begins with behaviorism, the view that mental concepts do not refer to anything inside us and that each type of mental state can be identified with a disposition of the organism to behave observably in certain ways under certain physical conditions. When this view began to look too much like a bald denial of the existence of the mind, some philosophers put forward identity theories, according to which mental processes are identical with brain processes in the same way that light is identical with electromagnetic radiation. But identity theories were left with the problem of explaining in nonmental terms what it means to say of a particular brain process that it is a thought or a sensation. After all, this can’t mean only that it is a certain kind of neurophysiological process. And given the aim of these theories, it couldn’t mean that the brain process has some mental effect. The proposed solution was a revival of behaviorism in a new form: thirst, for example, was identified not with a disposition to drink, but with a brain state, but that particular brain state’s being identical with thirst was now said to consist simply in the fact that it was typically

Another reductionist strategy, which I haven’t the space to discuss here, is to substitute for a theory of what mental states are a theory of the externally observable grounds on which we ascribe mental states to people, and to claim that this system of “asertibility conditions” is all the analysis the concepts need. One doesn’t identify mental phenomena with anything physical, because one doesn’t identify them with anything. But the conditions of applicability of mental concepts are, on this view, compatible with the world’s being nothing but a material system. This is essentially Daniel Dennett’s strategy in Consciousness Explained (Little, Brown, 1991) [discussed in Chapter 8 in this volume].

But see Geoffrey Madell, Mind and Materialism (Edinburgh University Press, 1988).
caused by dehydration and that it typically caused a disposition to drink. In this way it was thought that the identification of mental states with brain states could avoid all reference to nonphysical features.

These "causal behaviorist" analyses were eventually succeeded by a more technical theory called "functionalism," according to which mental concepts cannot be linked to behavior and circumstances individually but only as part of a single interconnected network. The behavior caused by thirst, for example, depends on the rest of a person's mental condition—his beliefs about where water is to be found and whether it is safe to drink, the strength of his desires to live or die, and so forth. Each mental state is a part of an integrated system that controls the organism's interaction with its environment; it is only by analyzing the role played by such states as thirst, pain, other kinds of sensation, belief, emotion, and desire, within the total system, that one can accurately describe their connection to behavior and external circumstances. Such a system may still permit mental states to be identified with brain states, provided the latter have causal or functional roles of the kind specified by the theory (still to be constructed) of how the integrated system works. Finally, functionalism led to what Searle calls Strong AI (Strong Artificial Intelligence)—the identification of mental states with computational states of a computer program that controls the organism's behavior—a program that is physically realized in the hardware (or wetware) of the brain.

All these theories attempt to reduce the mind to one or another aspect of a world that can be fully described by physics—the world of particles and fields. They have not been worked out in detail; they are just hopeful descriptions of the kind of thing a theory of the mind would have to be, together with some extremely sketchy examples. While each new proposal has been criticized by defenders of alternative reductionist accounts, Searle argues that there is one big thing wrong with all of them: they leave out consciousness.

No theory that leaves out consciousness can claim to be a theory of the mind, and no analysis of consciousness in nonmental terms is possible; therefore, no materialistic reduction of the mental can succeed. Searle contends that none of these theories could possibly provide an account of what pain, hunger, belief, vision, and the like really are, because all they talk about is what is externally observable—the organism's behavior and its causal structure—and a description of something exclusively in those terms fails to guarantee that it has any consciousness at all: each of these behaviorist, functionalist, or computational theories could be satisfied by an unconscious zombie of sufficient physical complexity.

The crucial question is not "Under what conditions would we attribute mental states to other people?" but rather, "What is it that people actually have when they have mental states?" "What are mental phenomena?" as distinct from "How do we find out about them and how do they function causally in the life of the organism?" (p. 23)

We attribute consciousness to other people and to animals on the basis of their behavior, but this is simply evidence of consciousness rather than proof of it, and it has to be supplemented by evidence of physiological similarity: since we believe in the uniformity of nature, we naturally infer that creatures who behave similarly to us and have sense organs and nervous systems physically similar to ours also have conscious experiences of some kind. But, Searle argues, no quantity of facts about physical behavior or functional organization by themselves entail that a system is conscious at all—and any theory that claims, for example, that vision is "nothing but" a certain state of the organism must, to be adequate, have the consequence that if the organism is in that state, it can't fail to be conscious. Otherwise, it will leave out the most important thing about vision and, whatever its other merits, won't qualify as an account of what vision is.

Not only do materialist reductions fail to imply that the system is conscious; it is clear in advance that no further development along the same lines, no added structural or behavioral complications, could do so. The reason is that there is a crucial difference between conscious phenomena and behavioral or physiological phenomena that makes the former irreducible to the latter: consciousness is, in Searle's terms, "ontologically subjective." That is, its essential features cannot be described entirely from an external, third-person point of view. Even the physiological description of what goes on inside the skull is external in this sense: it is described from outside. It is not enough to summarize the third-person observations, behavioral or physiological, that lead us to ascribe conscious mental states to others. The first-person point of view, which reveals what a conscious mental state is like for its subject, is indispensable.

This becomes clear when we ask, What is consciousness? Though we can describe certain of its features and identify more specific types of mental
phenomena as instances, it is so basic that it can't be defined in terms of anything else. You, reader, are conscious at this very moment, and your conscious condition includes such things as the way this page looks to you; the feel of the paper between your fingers; the shirt on your back; and the chair on which you're sitting; the sounds you hear of music or surf or police sirens in the background; and your experience of reading this sentence. Searle's claim is that no amount of third-person analysis, whether behavioral, causal, or functional, could possibly tell us what these experiences are in themselves—what they consist of, as distinguished from their causes and effects. This is perfectly obvious because subjective facts about what it's like for someone to be in a certain condition—what it's like from his point of view—can't be identified with facts about how things are, not from anyone's point of view or for anyone, but just in themselves. Facts about your external behavior or the electrical activity or functional organization of your brain may be closely connected with your conscious experiences, but they are not facts about what it's like for you to hear a police siren. 

Searle believes that the persistence of materialistic reductionism in the face of its evident falsity requires explanation. He likens it to the constant repetition by a compulsive neurotic of the same destructive pattern of behavior, and he hopes that by bringing to light its underlying causes he can break the hold of the compulsion. It is evident, both from what they say and from what they do, that reductionists are convinced in advance that some materialist theory of the mind must be correct: they just have to find it. This assumption is part of a scientific worldview to which they can see no alternative. But underlying the assumption, according to Searle, are two crucial misconceptions. The first is that we have to choose between materialism and dualism:

What I want to insist on, ceaselessly, is that one can accept the obvious facts of physics—for example, that the world is made up entirely of physical particles in fields of force—without at the same time denying the obvious facts about our own experiences—for example, that we are all conscious and that our conscious states have quite specific irreducible phenomenological [that is, subjective] properties. The mistake is to suppose that these two theses are inconsistent, and that mistake derives from accepting the presuppositions behind the traditional vocabulary. My view is emphatically not a form of dualism. I reject both property and substance dualism; but precisely for the reasons that I reject dualism, I reject materialism and monism as well. The deep mistake is to suppose that one must choose between these views.

Once you accept our worldview the only obstacle to granting consciousness its status as a biological feature of organisms is the outmoded dualistic/materialistic assumption that the "mental" character of consciousness makes it impossible for it to be a "physical" property. (pp. 28, 91)

This radical thesis, that consciousness is a physical property of the brain in spite of its subjectivity, and that it is irreducible to any other physical properties, is the metaphysical heart of Searle's position. The point here, however, is that Searle contends that materialists are drawn to implausible forms of psychophysical reduction because they assume that if mental states cannot be explained in such terms, then the inescapable alternative is dualism: they would then have to admit that nonphysical substances or properties are basic features of reality. And the fear of dualism, with its religious and spiritualist and otherwise unscientific associations, drives them to embrace reductionist materialism at any intellectual cost: "Materialism is thus in a sense the finest flower of dualism."

To escape from this bind, says Searle, we have to free ourselves of the urge to ask whether there are one or two ultimate kinds of things and properties. We should not start counting in the first place.

He is absolutely right about the fear of dualism (indeed, I believe he himself is not immune to its effects). Its most bizarre manifestation is yet another theory, called "eliminative" materialism. This is the view that, because mental states can't be accommodated within the world described by physics, they don't exist—just as witches and ghosts don't exist. They can be dismissed as postulates of a primitive theory customarily referred to as "folk psychology"—about which P. F. Strawson has remarked that it is the province of "such simple folk as Shakespeare, Tolstoy, Proust, and Henry James." Searle patiently pulverizes this view, but his real point is that the entire materialist tradition is in truth eliminative: all materialist theories deny the reality of the mind, but most of them disguise the fact (from themselves as well as from others) by identifying the mind with something else.

The second crucial misconception behind the compulsive search for materialist theories, according to Searle, is a simple but enormously destructive mistake about objectivity:

There is a persistent confusion between the claim that we should try as much as possible to eliminate personal subjective prejudices from the search

---

for truth and the claim that the real world contains no elements that are irreducibly subjective. And this confusion in turn is based on a confusion between the epistemological sense of the subjective/objective distinction, and the ontological sense. Epistemically, the distinction marks different degrees of independence of claims from the vagaries of special values, personal prejudices, points of view, and emotions. Ontologically, the distinction marks different categories of empirical reality. (p. 19)

This seems to me entirely convincing, and very important. Science must of course strive for epistemic objectivity—objective knowledge—by using methods that compensate for differences in points of view and that permit different observers to arrive at the same conception of what is the case. But it is a gross confusion to conclude from this that nothing that has or includes a point of view can be an object of scientific investigation. Subjective points of view are themselves parts of the real world, and if they and their properties are to be described adequately, their ontologically subjective character—the subjectivity of their nature—must be acknowledged. Furthermore, this can be done, in the epistemic sense, objectively: although only you are now experiencing the look of the page in front of you, others can know that you are, and can know a good deal about what that experience is like for you. It is an objective truth that you are now having a certain subjective visual experience.

If we accept this distinction, the question becomes, How can we form an epistemically objective scientific conception of a world that contains not only the familiar ontologically objective facts described by physics, chemistry, and biology, but also the ontologically subjective facts of consciousness? And that brings us, finally, to Searle’s own view, which he calls “biological naturalism,” and which combines acceptance of the irreducible subjectivity of the mental with rejection of the dichotomy between mental and physical: “Consciousness . . . is a biological feature of human and certain animal brains. It is caused by neurobiological processes and is as much part of the natural biological order as any other biological features such as photosynthesis, digestion, or mitosis” (p. 90).

And in spite of his antireductionism, he also writes as follows:

Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense of “higher-level” or “emergent” in which solidity is a higher-level emergent property of H₂O molecules when they are in a lattice structure (ice), and liquidity is similarly a higher-level emergent property of H₂O molecules when they are, roughly speaking, rolling around on each other (water). Consciousness is a mental, and therefore physical, property of the brain in the sense in which liquidity is a property of systems of molecules. (p. 14)

If this view could be clarified in a way that distinguished it from the alternatives, it would be a major addition to the possible answers to the mind-body problem. But I don’t think it can be.

Suppose we grant that states of consciousness are properties of the brain caused by, but not reducible to, its neural activity. This means that your brain, for instance, has a point of view of which all your current experiences are aspects. But what is the justification for calling these irreducibly subjective features of the brain physical? What would it even mean to call them physical? Certainly they are “higher-order” in the sense that they can be ascribed only to the system as a whole and not to its microscopic parts; they are also “emergent” in the sense of being explained only by the causal interactions of those parts. But however great the variety of physical phenomena may be, ontological objectivity is one of their central defining conditions; and as we have seen, Searle insists that consciousness is ontologically subjective.

Searle doesn’t say enough about this question. Perhaps he believes that if brains are made up of physical particles, it follows automatically that all their properties are physical. And he quotes a remark of Noam Chomsky that as soon as we come to understand anything, we call it “physical.” But if “physical” is in this sense merely an honorific term (another way I’ve heard Chomsky put the point), what is the metaphysical content of Searle’s claim that mental properties are physical, and his emphatic rejection of property dualism? He says, after all, that the ontological distinction between subjective and objective marks “different categories of empirical reality.” To say further that we are “left with a universe that contains an irreducibly subjective physical component as a component of physical reality” merely couches an essentially dualistic claim in language that expresses a strong aversion to dualism.

Perhaps we could adopt Searle’s use of the word “physical,” but the basic issue is more than verbal. It is the issue of how to construct an intelligible and complete scientific worldview once we deny the reducibility of the mental to the nonmental. As Searle points out, we cannot do so by continuing on the path that physical science has followed since the seventeenth century because that depended on excluding the mind of the observer from the world being observed and described. To propose that conscious-

---

6Searle identifies me as a defender of property dualism. I prefer the term “dual aspect theory,” to express the view deriving from Spinoza that mental phenomena are the subjective aspects of states that can also be described physically. But all I would claim for the idea is that it is somewhat less unacceptable than the other unacceptable theories currently on offer. I share Searle’s aversion to both dualism and materialism, and believe a solution to the mind-body problem is nowhere in sight.
ness is an intrinsic subjective property of the brain caused by its neural activity is the first step on a different path—the right one, in my opinion. But there are large problems ahead, and they are not just empirical but philosophical.

Even if we learn a great deal more than we know now about the physiological causes of consciousness, it will not, as Searle is aware, make the relation of consciousness to the behavior of neurons analogous to the relation of liquidity to the behavior of H₂O molecules. In the latter case the relation is transparent: we can see how liquidity is the logical result of the molecules "rolling around on each other" at the microscopic level. Nothing comparable to be expected in the case of neurons, even though it is empirically evident that states of consciousness are the necessary consequences of neuronal activity. Searle has an interesting discussion of this difference, which he says results only from a limitation of our powers of conception: we can represent the necessary relation between the macro and micro levels of water since we picture them both from the outside, but we can't do this with subjectivity, which we have to imagine from the inside, whether it is ours or someone else's. I agree, but I believe this means we do not really understand the claim that mental states are states of the brain: we are still unable to form a conception of how consciousness arises in matter, even if we are certain that it does.⁷

III

Searle's second set of arguments against the computer model of mind depends on the specific nature of computers and is more distinctively Searle's own: it grows out of his long-standing concern with the theory of meaning and the "intentionality" of mental states— their capacity to mean something or refer to something outside themselves, and their consequent susceptibility to judgments of truth or falsity, correctness or incorrectness.⁸

How is it possible for computers to answer the questions we put to them—arithmetic questions, for example? The explanation has two parts. First, it is possible to formulate each of those questions by using a string of symbols—letters or numerals—selected from a short list and distinguished by their shapes, and to devise a finite set of rules for manipulating those symbols, which has the following property: if you start with the string corresponding to the question, and follow the rules for moving, removing, and adding symbols, you will arrive, after a finite series of uniquely determined steps, at a point where the rules tell you to stop; and the last string you have produced will correspond to the answer.⁹ you only have to read it. But to follow the rules for manipulating the symbols, you don't have to know what they mean, or whether they mean anything: you just have to identify their shapes.

Such rules are called rules of syntax (as opposed to rules of semantics, which you need to interpret a string as meaning something). The beauty of it is that we could train someone to do long division, for example, completely mechanically, using a set of rules and tables, without his knowing that the symbols he was writing down represented numbers or anything at all; but if he followed the syntactic rules, he would come up with what we (but not he) could read as the answer.¹⁰

The second part of the explanation is that there are, besides writing on paper, different ways of encoding arithmetic in a set of symbols and syntactic rules, and, for some of those ways, it is possible to design physical machines to carry out mechanically all the steps in juggling the symbols that the rules prescribe. Instead of a person following the syntactic rules mechanically without knowing what the symbols mean, a physical mechanism can carry out the same operations on the symbols automatically. But note, not only does this mechanism not know what the symbols mean, it doesn't even know, as our semantically deprived scribe did, that it is following rules for their manipulation, syntactic rules. It doesn't know anything—in fact, it isn't following rules at all, but just functioning in accordance with the laws of physics, in ways that clever engineers have designed to permit us to interpret the results both syntactically and semantically.

Searle's well-known "Chinese room" argument described a conscious person who, without knowing Chinese, follows rules for manipulating Chinese characters and puts together sentences intelligible to people who know Chinese. The point being made against the computer model of mind was that syntax alone can't yield semantics.¹¹ In *The Rediscovery of the Mind*, he extends the argument to show that physics alone can't yield syntax. Following rules, even purely syntactic rules, is an irreducibly men-

---

⁷ Colin McGinn, in *The Problem of Consciousness* (Blackwell, 1991), suggests that we are constitutionally incapable of arriving at such an understanding. Though he could be right, I think his pessimism is premature.


⁹ Such a procedure is called an "algorithm."

¹⁰ The general theory of this type of computability was developed by Alan Turing.

¹¹ See his "Minds, Brains, and Programs."
tal process—an "intentional" process in which the meaning of the rules themselves must be grasped by a conscious mind. It is not just a matter of regularity in physical behavior:

A physical state of a system is a computational state only relative to the assignment to that state of some computational role, function, or interpretation . . . notions such as computation, algorithm, and program do not name intrinsic physical features of systems: Computational states are not discovered within the physics, they are assigned to the physics. . . .

The aim of natural science is to discover and characterize features that are intrinsic to the natural world. By its own definitions of computation and cognition, there is no way that computational cognitive science could ever be a natural science, because computation is not an intrinsic feature of the world. It is assigned relative to observers. (pp. 210, 212; italics in original)

Searle's distinction between what is intrinsic to the thing observed and what is relative to an observer or interpreter is a fundamental one. He argues that intrinsic intentionality—that is, the capacity for grasping the meaning of statements and consciously following rules—occurs only in minds. Words on a page or electrical resistances in a computer chip can be said to mean something, or to obey rules of grammar or arithmetic, only in the derivative sense that our minds can interpret them that way, in virtue of their arrangement. This means that the claim that the brain is a computer would imply that the brain has intentionality and follows rules of computation not intrinsically but only relative to the interpretation of its user. And who is the user supposed to be? If the brain is a computer, it does not have intrinsic intentionality. If it has intrinsic intentionality, it must be more than a computer. Searle chooses the second alternative. He also argues that those theories which try to construe the brain as a computer always surreptitiously assume a mind or "homunculus" as its interpreter.

There is a lot more to this argument, and though I find its negative conclusions persuasive, questions I have not touched on could be raised about Searle's positive theory of intrinsic intentionality, which is meant to be consistent with his biological naturalism. As with consciousness, it remains extremely difficult to see how intrinsic intentionality could be a property of the physical organism. But instead of pursuing these questions here, I will turn to Searle's views about the unconscious and its relation to consciousness, for these serve to bring together the two parts of the argument.

Searle has put great weight on the claim that subjective consciousness is not reducible to anything else. But most of our mental states are not conscious. Take all the beliefs and hopes and intentions you may have but are not thinking about right now—the belief that there's a leaning tower in Pisa, for example. It just became conscious, but you've probably believed it for years. If such beliefs can exist unconsciously, then consciousness is not an essential feature of mental life, and it must be possible for intentional mental states to be embodied in a purely material brain. So it could be argued that even for those mental states which are conscious, their subjective, experienced character is not essential for their intentionality. Perhaps consciousness is just a kind of subjective "tint" that sometimes gets added to the truly functional black and white of mental states so that a theory of mind could dismiss consciousness as inessential.

Here is Searle's reply to the suggestion:

We understand the notion of an unconscious mental state only as a possible content of consciousness, only as the sort of thing that, though not conscious, and perhaps impossible to bring to consciousness for various reasons [such as repression], is nonetheless the sort of thing that could be or could have been conscious. (p. 156)

He calls this the "connection principle"; his argument for it is that even unconscious mental states must have a distinctively subjective character to qualify as beliefs, thoughts, desires, and so on. To be about anything, and therefore true or false, right or wrong, a state must belong at least potentially to the point of view of some subject. Searle acknowledges that it was a neurophysiological state of your brain that made it true two hours ago that you believed that there was a leaning tower in Pisa, but he argues that neurophysiology alone cannot qualify that state as a belief, because no physiological description by itself implies that the brain state has any intentionality or meaning at all—even if we add to the description on account of the physical behavior it might cause. His conclusion is that neurophysiological states can be assigned intentionality only derivatively, in virtue of the conscious states they are capable of producing: "The ontology of the unconscious consists in subjective features of the brain capable of causing subjective conscious thoughts." This has the surprising consequence that a deep, allegedly psychological mechanism like Chomsky's Language Acquisition Device, which allows a child to learn the grammar of a language on the basis of the samples of speech it encounters at an early age, is not a set of unconscious mental rules at all, but simply a physical mechanism—for it is incapable of giving rise to subjective conscious thought whose content consists of those rules themselves. So in Searle's view, the child's conformity to the rules in learning language is not an example of intrinsic intentionality, but only intentionality assigned by the linguist-observer.
In sum, consciousness is the essence of the mental, even if most mental states are not conscious at any given time. One cannot study any aspect of mental experience without including it or its possibility in the definition of what one is trying to understand. In particular, intentionality is inseparable from it.

The Rediscovery of the Mind is trenchant, aggressive, and beautifully clear, in Searle's best "What is all this nonsense?" style. It is a valuable antidote to one of the dominant illusions of our age.

II

ETHICS AND POLITICAL PHILOSOPHY
OTHER MINDS
CRITICAL ESSAYS 1969–1994

THOMAS NAGEL