

The Mind-Body Problem

Could calculating machines have pains, Martians have expectations and disembodied spirits have thoughts? The modern functionalist approach to psychology raises the logical possibility that they could

by Jerry A. Fodor

**NOTICE: THIS MATERIAL MAY
BE PROTECTED BY COPYRIGHT
LAW (TITLE 17, U.S. CODE)**

Modern philosophy of science has been devoted largely to the formal and systematic description of the successful practices of working scientists. The philosopher does not try to dictate how scientific inquiry and argument ought to be conducted. Instead he tries to enumerate the principles and practices that have contributed to good science. The philosopher has devoted the most attention to analyzing the methodological peculiarities of the physical sciences. The analysis has helped to clarify the nature of confirmation, the logical structure of scientific theories, the formal properties of statements that express laws and the question of whether theoretical entities actually exist.

It is only rather recently that philosophers have become seriously interested in the methodological tenets of psychology. Psychological explanations of behavior refer liberally to the mind and to states, operations and processes of the mind. The philosophical difficulty comes in stating in unambiguous language what such references imply.

Traditional philosophies of mind can be divided into two broad categories: dualist theories and materialist theories. In the dualist approach the mind is a nonphysical substance. In materialist theories the mental is not distinct from the physical; indeed, all mental states, properties, processes and operations are in principle identical with physical states, properties, processes and operations. Some materialists, known as behaviorists, maintain that all talk of mental causes can be eliminated from the language of psychology in favor of talk of environmental stimuli and behavioral responses. Other materialists, the identity theorists, contend that there are mental causes and that they are identical with neurophysiological events in the brain.

In the past 15 years a philosophy of mind called functionalism that is neither dualist nor materialist has emerged from philosophical reflection on developments in artificial intelligence, computational theory, linguistics, cybernet-

ics and psychology. All these fields, which are collectively known as the cognitive sciences, have in common a certain level of abstraction and a concern with systems that process information. Functionalism, which seeks to provide a philosophical account of this level of abstraction, recognizes the possibility that systems as diverse as human beings, calculating machines and disembodied spirits could all have mental states. In the functionalist view the psychology of a system depends not on the stuff it is made of (living cells, metal or spiritual energy) but on how the stuff is put together. Functionalism is a difficult concept, and one way of coming to grips with it is to review the deficiencies of the dualist and materialist philosophies of mind it aims to displace.

The chief drawback of dualism is its failure to account adequately for mental causation. If the mind is nonphysical, it has no position in physical space. How, then, can a mental cause give rise to a behavioral effect that has a position in space? To put it another way, how can the nonphysical give rise to the physical without violating the laws of the conservation of mass, of energy and of momentum?

The dualist might respond that the problem of how an immaterial substance can cause physical events is not much obscurer than the problem of how one physical event can cause another. Yet there is an important difference: there are many clear cases of physical causation but not one clear case of nonphysical causation. Physical interaction is something philosophers, like all other people, have to live with. Nonphysical interaction, however, may be no more than an artifact of the immaterialist construal of the mental. Most philosophers now agree that no argument has successfully demonstrated why mind-body causation should not be regarded as a species of physical causation.

Dualism is also incompatible with the practices of working psychologists. The psychologist frequently applies the experimental methods of the physical sci-

ences to the study of the mind. If mental processes were different in kind from physical processes, there would be no reason to expect these methods to work in the realm of the mental. In order to justify their experimental methods many psychologists urgently sought an alternative to dualism.

In the 1920's John B. Watson of Johns Hopkins University made the radical suggestion that behavior does not have mental causes. He regarded the behavior of an organism as its observable responses to stimuli, which he took to be the causes of its behavior. Over the next 30 years psychologists such as B. F. Skinner of Harvard University developed Watson's ideas into an elaborate world view in which the role of psychology was to catalogue the laws that determine causal relations between stimuli and responses. In this "radical behaviorist" view the problem of explaining the nature of the mind-body interaction vanishes; there is no such interaction.

Radical behaviorism has always worn an air of paradox. For better or worse, the idea of mental causation is deeply ingrained in our everyday language and in our ways of understanding our fellow men and ourselves. For example, people commonly attribute behavior to beliefs, to knowledge and to expectations. Brown puts gas in his tank because he believes the car will not run without it. Jones writes not "acheive" but "achieve" because he knows the rule about putting *i* before *e*. Even when a behavioral response is closely tied to an environmental stimulus, mental processes often intervene. Smith carries an umbrella because the sky is cloudy, but the weather is only part of the story. There are apparently also mental links in the causal chain: observation and expectation. The clouds affect Smith's behavior only because he observes them and because they induce in him an expectation of rain.

The radical behaviorist is unmoved by appeals to such cases. He is prepared to dismiss references to mental causes, however plausible they may seem, as the residue of outworn creeds. The radical

behaviorist predicts that as psychologists come to understand more about the relations between stimuli and responses they will find it increasingly possible to explain behavior without postulating mental causes.

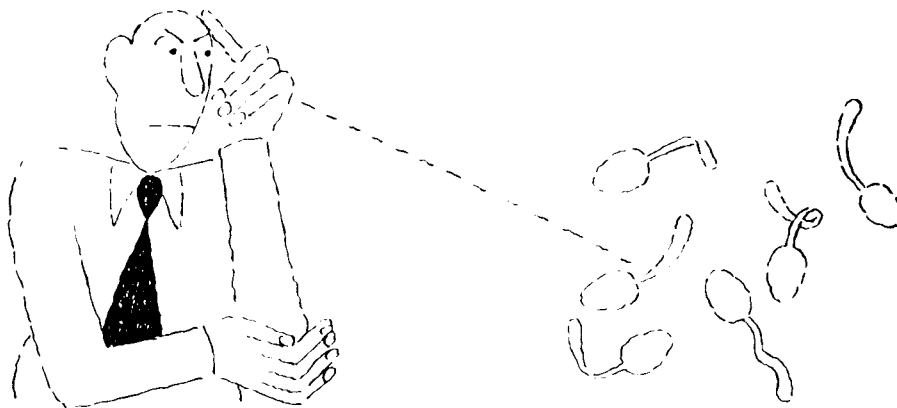
The strongest argument against behaviorism is that psychology has not turned out this way; the opposite has happened. As psychology has matured, the framework of mental states and processes that is apparently needed to account for experimental observations has grown all the more elaborate. Particularly in the case of human behavior psychological theories satisfying the methodological tenets of radical behaviorism have proved largely sterile, as would be expected if the postulated mental processes are real and causally effective.

Nevertheless, many philosophers were initially drawn to radical behaviorism because, paradoxes and all, it seemed better than dualism. Since a psychology committed to immaterial substances was unacceptable, philosophers turned to radical behaviorism because it seemed to be the only alternative materialist philosophy of mind. The choice, as they saw it, was between radical behaviorism and ghosts.

By the early 1960's philosophers began to have doubts that dualism and radical behaviorism exhausted the possible approaches to the philosophy of mind. Since the two theories seemed unattractive, the right strategy might be to develop a materialist philosophy of mind that nonetheless allowed for mental causes. Two such philosophies emerged, one called logical behaviorism and the other called the central-state identity theory.

Logical behaviorism is a semantic theory about what mental terms mean. The basic idea is that attributing a mental state (say thirsty) to an organism is the same as saying that the organism is disposed to behave in a particular way (for example to drink if there is water available). On this view every mental ascription is equivalent in meaning to an if-then statement (called a behavioral hypothetical) that expresses a behavioral disposition. For example, "Smith is thirsty" might be taken to be equivalent to the dispositional statement "If there were water available, then Smith would drink some." By definition a behavioral hypothetical includes no mental terms. The if-clause of the hypothetical speaks only of stimuli and the then-clause speaks only of behavioral responses. Since stimuli and responses are physical events, logical behaviorism is a species of materialism.

The strength of logical behaviorism is that by translating mental language into the language of stimuli and responses it provides an interpretation of psychological explanations in which behavioral



DUALISM is the philosophy of mind that regards the mind as a nonphysical substance. It divides everything there is in the world into two distinct categories: the mental and the physical. The chief difficulty with dualism is its failure to account adequately for the causal interaction of the mental and the physical. It is not evident how a nonphysical mind could give rise to any physical effects without violating the laws of conservation of mass, energy and momentum.

effects are attributed to mental causes. Mental causation is simply the manifestation of a behavioral disposition. More precisely, mental causation is what happens when an organism has a behavioral disposition and the if-clause of the behavioral hypothetical expressing the disposition happens to be true. For example, the causal statement "Smith drank some water because he was thirsty" might be taken to mean "If there were

water available, then Smith would drink some, and there was water available."

I have somewhat oversimplified logical behaviorism by assuming that each mental ascription can be translated by a unique behavioral hypothetical. Actually the logical behaviorist often maintains that it takes an open-ended set (perhaps an infinite set) of behavioral hypotheticals to spell out the behavioral disposition expressed by a mental term.



RADICAL BEHAVIORISM is the philosophy of mind that denies the existence of the mind and mental states, properties, processes and operations. The radical behaviorist believes behavior does not have mental causes. He considers the behavior of an organism to be its responses to stimuli. The role of psychology is to catalogue the relations between stimuli and responses.

The mental ascription "Smith is thirsty" might also be satisfied by the hypothetical "If there were orange juice available, then Smith would drink some" and by a host of other hypotheticals. In any event the logical behaviorist does not usually maintain he can actually enumerate all the hypotheticals that correspond to a behavioral disposition expressing a given mental term. He only insists that in principle the meaning of any mental term can be conveyed by behavioral hypotheticals.

The way the logical behaviorist has interpreted a mental term such as thirsty is modeled after the way many philosophers have interpreted a physical disposition such as fragility. The physical disposition "The glass is fragile" is often taken to mean something like "If the glass were struck, then it would break." By the same token the logical behaviorist's analysis of mental causation is similar to the received analysis of one kind of physical causation. The causal statement "The glass broke because it was fragile" is taken to mean something like "If the glass were struck, then it would break, and the glass was struck."

By equating mental terms with behav-

ioral dispositions the logical behaviorist has put mental terms on a par with the nonbehavioral dispositions of the physical sciences. That is a promising move, because the analysis of nonbehavioral dispositions is on relatively solid philosophical ground. An explanation attributing the breaking of a glass to its fragility is surely something even the staunchest materialist can accept. By arguing that mental terms are synonymous with dispositional terms, the logical behaviorist has provided something the radical behaviorist could not: a materialist account of mental causation.

Nevertheless, the analogy between mental causation as construed by the logical behaviorist and physical causation goes only so far. The logical behaviorist treats the manifestation of a disposition as the sole form of mental causation, whereas the physical sciences recognize additional kinds of causation. There is the kind of causation where one physical event causes another, as when the breaking of a glass is attributed to its having been struck. In fact, explanations that involve event-event causation are presumably more basic than dispositional explanations, because the manifestation of a disposition (the breaking of a fragile glass) always involves event-

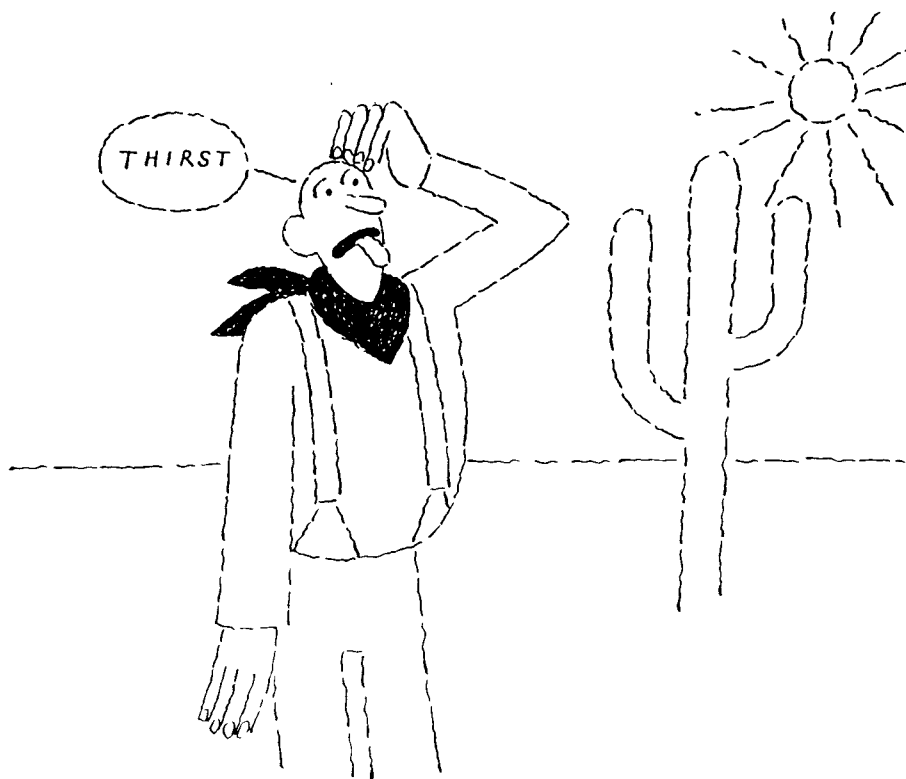
event causation and not vice versa. In the realm of the mental many examples of event-event causation involve one mental state's causing another, and for this kind of causation logical behaviorism provides no analysis. As a result the logical behaviorist is committed to the tacit and implausible assumption that psychology requires a less robust notion of causation than the physical sciences require.

Event-event causation actually seems to be quite common in the realm of the mental. Mental causes typically give rise to behavioral effects by virtue of their interaction with other mental causes. For example, having a headache causes a disposition to take aspirin only if one also has the desire to get rid of the headache, the belief that aspirin exists, the belief that taking aspirin reduces headaches and so on. Since mental states interact in generating behavior, it will be necessary to find a construal of psychological explanations that posits mental processes: causal sequences of mental events. It is this construal that logical behaviorism fails to provide.

Such considerations bring out a fundamental way in which logical behaviorism is quite similar to radical behaviorism. It is true that the logical behaviorist, unlike the radical behaviorist, acknowledges the existence of mental states. Yet since the underlying tenet of logical behaviorism is that references to mental states can be translated out of psychological explanations by employing behavioral hypotheticals, all talk of mental states and processes is in a sense heuristic. The only facts to which the behaviorist is actually committed are facts about relations between stimuli and responses. In this respect logical behaviorism is just radical behaviorism in a semantic form. Although the former theory offers a construal of mental causation, the construal is Pickwickian. What does not really exist cannot cause anything, and the logical behaviorist, like the radical behaviorist, believes deep down that mental causes do not exist.

An alternative materialist theory of the mind to logical behaviorism is the central-state identity theory. According to this theory, mental events, states and processes are identical with neurophysiological events in the brain, and the property of being in a certain mental state (such as having a headache or believing it will rain) is identical with the property of being in a certain neurophysiological state. On this basis it is easy to make sense of the idea that a behavioral effect might sometimes have a chain of mental causes; that will be the case whenever a behavioral effect is contingent on the appropriate sequence of neurophysiological events.

The central-state identity theory acknowledges that it is possible for mental



LOGICAL BEHAVIORISM is a semantic thesis about what mental terms mean. The logical behaviorist maintains that mental terms express behavioral dispositions. Consider the mental state of being thirsty. The logical behaviorist maintains that the sentence "Smith is thirsty" might be taken as equivalent in meaning to the dispositional statement "If there were water available, then Smith would drink some." The strength of logical behaviorism is that it provides an account of mental causation: the realization of a behavioral disposition. For example, the causal statement "Smith drank some water because he was thirsty" might be taken to mean "If there were water available, then Smith would drink some, and there was water available."



CENTRAL-STATE IDENTITY THEORY is the philosophy of mind that equates mental events, states and processes with neuro-

physiological events. Property of being in a given mental state is identical with the property of being in a given neurophysiological state.

causes to interact causally without ever giving rise to any behavioral effect, as when a person thinks for a while about what he ought to do and then decides to do nothing. If mental processes are neurophysiological, they must have the causal properties of neurophysiological processes. Since neurophysiological processes are presumably physical processes, the central-state identity theory ensures that the concept of mental causation is as rich as the concept of physical causation.

The central-state identity theory provides a satisfactory account of what the mental terms in psychological explanations refer to, and so it is favored by psychologists who are dissatisfied with behaviorism. The behaviorist maintains that mental terms refer to nothing or that they refer to the parameters of stimulus-response relations. Either way the existence of mental entities is only illusory. The identity theorist, on the other hand, argues that mental terms refer to neurophysiological states. Thus he can take seriously the project of explaining behavior by appealing to its mental causes.

The chief advantage of the identity theory is that it takes the explanatory constructs of psychology at face value, which is surely something a philosophy of mind ought to do if it can. The identi-

ty theory shows how the mentalistic explanations of psychology could be not mere heuristics but literal accounts of the causal history of behavior. Moreover, since the identity theory is not a semantic thesis, it is immune to many arguments that cast in doubt logical behaviorism. A drawback of logical behaviorism is that the observation "John has a headache" does not seem to mean the same thing as a statement of the form "John is disposed to behave in such and such a way." The identity theorist, however, can live with the fact that "John has a headache" and "John is in such and such a brain state" are not synonymous. The assertion of the identity theorist is not that these sentences mean the same thing but only that they are rendered true (or false) by the same neurophysiological phenomena.

The identity theory can be held either as a doctrine about mental particulars (John's current pain or Bill's fear of animals) or as a doctrine about mental universals, or properties (having a pain or being afraid of animals). The two doctrines, called respectively token physicalism and type physicalism, differ in strength and plausibility. Token physicalism maintains only that all the mental particulars that happen to exist are neurophysiological, whereas type physicalism makes the more sweeping asser-

tion that all the mental particulars there could possibly be are neurophysiological. Token physicalism does not rule out the logical possibility of machines and disembodied spirits having mental properties. Type physicalism dismisses this possibility because neither machines nor disembodied spirits have neurons.

Type physicalism is not a plausible doctrine about mental properties even if token physicalism is right about mental particulars. The problem with type physicalism is that the psychological constitution of a system seems to depend not on its hardware, or physical composition, but on its software, or program. Why should the philosopher dismiss the possibility that silicon-based Martians have pains, assuming that the silicon is properly organized? And why should the philosopher rule out the possibility of machines having beliefs, assuming that the machines are correctly programmed? If it is logically possible that Martians and machines could have mental properties, then mental properties and neurophysiological processes cannot be identical, however much they may prove to be coextensive.

What it all comes down to is that there seems to be a level of abstraction at which the generalizations of psychology are most naturally pitched. This level of

abstraction cuts across differences in the physical composition of the systems to which psychological generalizations apply. In the cognitive sciences, at least, the natural domain for psychological theorizing seems to be all systems that process information. The problem with type physicalism is that there are possible information-processing systems with the same psychological constitution as human beings but not the same physical organization. In principle all kinds of physically different things could have human software.

This situation calls for a relational account of mental properties that abstracts them from the physical structure of their bearers. In spite of the objections to logical behaviorism that I presented above, logical behaviorism was at least on the right track in offering a relational interpretation of mental properties: to have a headache is to be disposed to exhibit a certain pattern of relations between the stimuli one encounters and the responses one exhibits. If that is what having a headache is, how-

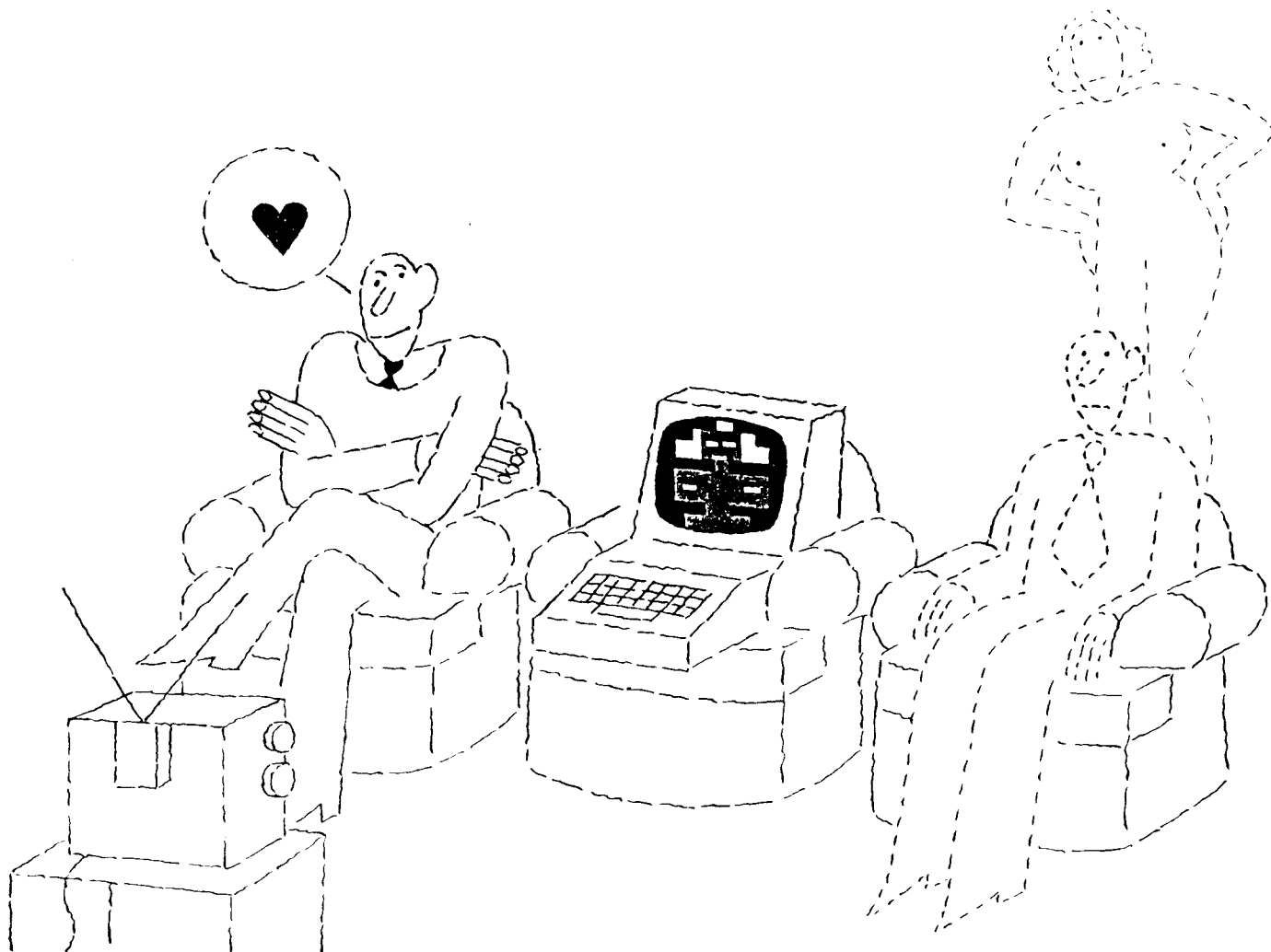
ever, there is no reason in principle why only heads that are physically similar to ours can ache. Indeed, according to logical behaviorism, it is a necessary truth that any system that has our stimulus-response contingencies also has our headaches.

All of this emerged 10 or 15 years ago as a nasty dilemma for the materialist program in the philosophy of mind. On the one hand the identity theorist (and not the logical behaviorist) had got right the causal character of the interactions of mind and body. On the other the logical behaviorist (and not the identity theorist) had got right the relational character of mental properties. Functionalism has apparently been able to resolve the dilemma. By stressing the distinction computer science draws between hardware and software the functionalist can make sense of both the causal and the relational character of the mental.

The intuition underlying functionalism is that what determines the psychological type to which a mental particular belongs is the causal role of the particu-

lar in the mental life of the organism. Functional individuation is differentiation with respect to causal role. A headache, for example, is identified with the type of mental state that among other things causes a disposition for taking aspirin in people who believe aspirin relieves a headache, causes a desire to rid oneself of the pain one is feeling, often causes someone who speaks English to say such things as "I have a headache" and is brought on by overwork, eye-strain and tension. This list is presumably not complete. More will be known about the nature of a headache as psychological and physiological research discovers more about its causal role.

Functionalism construes the concept of causal role in such a way that a mental state can be defined by its causal relations to other mental states. In this respect functionalism is completely different from logical behaviorism. Another major difference is that functionalism is not a reductionist thesis. It does not foresee, even in principle,



FUNCTIONALISM is the philosophy of mind based on the distinction that computer science draws between a system's hardware, or physical composition, and its software, or program. The psychology of a system such as a human being, a machine or a disembodied spirit

does not depend on the stuff the system is made of (neurons, diodes or spiritual energy) but on how that stuff is organized. Functionalism does not rule out the possibility, however remote it may be, of mechanical and ethereal systems having mental states and processes.

the elimination of mentalistic concepts from the explanatory apparatus of psychological theories.

The difference between functionalism and logical behaviorism is brought out by the fact that functionalism is fully compatible with token physicalism. The functionalist would not be disturbed if brain events turn out to be the only things with the functional properties that define mental states. Indeed, most functionalists fully expect it will turn out that way.

Since functionalism recognizes that mental particulars may be physical, it is compatible with the idea that mental causation is a species of physical causation. In other words, functionalism tolerates the materialist solution to the mind-body problem provided by the central-state identity theory. It is possible for the functionalist to assert both that mental properties are typically defined in terms of their relations and that interactions of mind and body are typically causal in however robust a notion of causality is required by psychological explanations. The logical behaviorist can endorse only the first assertion and the type physicalist only the second. As a result functionalism seems to capture the best features of the materialist alternatives to dualism. It is no wonder that functionalism has become increasingly popular.

Machines provide good examples of two concepts that are central to functionalism: the concept that mental states are interdefined and the concept that they can be realized by many systems. The illustration on the next page contrasts a behavioristic Coke machine with a mentalistic one. Both machines dispense a Coke for 10 cents. (The price has not been affected by inflation.) The states of the machines are defined by reference to their causal roles, but only the machine on the left would satisfy the behaviorist. Its single state (S_0) is completely specified in terms of stimuli and responses. S_0 is the state a machine is in if, and only if, given a dime as the input, it dispenses a Coke as the output.

The machine on the right in the illustration has interdefined states (S_1 and S_2), which are characteristic of functionalism. S_1 is the state a machine is in if, and only if, (1) given a nickel, it dispenses nothing and proceeds to S_2 , and (2) given a dime, it dispenses a Coke and stays in S_1 . S_2 is the state a machine is in if, and only if, (1) given a nickel, it dispenses a Coke and proceeds to S_1 , and (2) given a dime, it dispenses a Coke and a nickel and proceeds to S_1 . What S_1 and S_2 jointly amount to is the machine's dispensing a Coke if it is given a dime, dispensing a Coke and a nickel if it is given a dime and a nickel and waiting to be given a second nickel if it has been given a first one.

Since S_1 and S_2 are each defined by hypothetical statements, they can be viewed as dispositions. Nevertheless, they are not behavioral dispositions because the consequences an input has for a machine in S_1 or S_2 are not specified solely in terms of the output of the machine. Rather, the consequences also involve the machine's internal states.

Nothing about the way I have described the behavioristic and mentalistic Coke machines puts constraints on what they could be made of. Any system whose states bore the proper relations to inputs, outputs and other states could be one of these machines. No doubt it is reasonable to expect such a system to be constructed out of such things as wheels, levers and diodes (token physicalism for Coke machines). Similarly, it is reasonable to expect that our minds may prove to be neurophysiological (token physicalism for human beings).

Nevertheless, the software description of a Coke machine does not logically require wheels, levers and diodes for its concrete realization. By the same token, the software description of the mind does not logically require neurons. As far as functionalism is concerned a Coke machine with states S_1 and S_2 could be made of ectoplasm, if there is such stuff and if its states have the right causal properties. Functionalism allows for the possibility of disembodied Coke machines in exactly the same way and to the same extent that it allows for the possibility of disembodied minds.

To say that S_1 and S_2 are interdefined and realizable by different kinds of hardware is not, of course, to say that a Coke machine has a mind. Although interdefinition and functional specification are typical features of mental states, they are clearly not sufficient for mentality. What more is required is a question to which I shall return below.

Some philosophers are suspicious of functionalism because it seems too easy. Since functionalism licenses the individuation of states by reference to their causal role, it appears to allow a trivial explanation of any observed event E , that is, it appears to postulate an E -causer. For example, what makes the valves in a machine open? Why, the operation of a valve opener. And what is a valve opener? Why, anything that has the functionally defined property of causing valves to open.

In psychology this kind of question-begging often takes the form of theories that in effect postulate homunculi with the selfsame intellectual capacities the theorist set out to explain. Such is the case when visual perception is explained by simply postulating psychological mechanisms that process visual information. The behaviorist has often charged the mentalist, sometimes justifiably, of mongering this kind of question-begging pseudo explanation. The

charge will have to be met if functionally defined mental states are to have a serious role in psychological theories.

The burden of the accusation is not untruth but triviality. There can be no doubt that it is a valve opener that opens valves, and it is likely that visual perception is mediated by the processing of visual information. The charge is that such putative functional explanations are mere platitudes. The functionalist can meet this objection by allowing functionally defined theoretical constructs only where mechanisms exist that can carry out the function and only where he has some notion of what such mechanisms might be like. One way of imposing this requirement is to identify the mental processes that psychology postulates with the operations of the restricted class of possible computers called Turing machines.

A Turing machine can be informally characterized as a mechanism with a finite number of program states. The inputs and outputs of the machine are written on a tape that is divided into squares each of which includes a symbol from a finite alphabet. The machine scans the tape one square at a time. It can erase the symbol on a scanned square and print a new one in its place. The machine can execute only the elementary mechanical operations of scanning, erasing, printing, moving the tape and changing state.

The program states of the Turing machine are defined solely in terms of the input symbols on the tape, the output symbols on the tape, the elementary operations and the other states of the program. Each program state is therefore functionally defined by the part it plays in the overall operation of the machine. Since the functional role of a state depends on the relation of the state to other states as well as to inputs and outputs, the relational character of the mental is captured by the Turing-machine version of functionalism. Since the definition of a program state never refers to the physical structure of the system running the program, the Turing-machine version of functionalism also captures the idea that the character of a mental state is independent of its physical realization. A human being, a roomful of people, a computer and a disembodied spirit would all be a Turing machine if they operated according to a Turing-machine program.

The proposal is to restrict the functional definition of psychological states to those that can be expressed in terms of the program states of Turing machines. If this restriction can be enforced, it provides a guarantee that psychological theories will be compatible with the demands of mechanisms. Since Turing machines are very simple devices, they are in principle quite easy to

build. Consequently by formulating a psychological explanation as a Turing-machine program the psychologist ensures that the explanation is mechanistic, even though the hardware realizing the mechanism is left open.

There are many kinds of computational mechanisms other than Turing machines, and so the formulation of a functionalist psychological theory in Turing-machine notation provides only a sufficient condition for the theory's being mechanically realizable. What makes the condition interesting, however, is that the simple Turing machine can perform many complex tasks. Although the elementary operations of the Turing machine are restricted, iterations of the operations enable the machine to carry out any well-defined computation on discrete symbols.

An important tendency in the cognitive sciences is to treat the mind chiefly as a device that manipulates symbols. If

a mental process can be functionally defined as an operation on symbols, there is a Turing machine capable of carrying out the computation and a variety of mechanisms for realizing the Turing machine. Where the manipulation of symbols is important the Turing machine provides a connection between functional explanation and mechanistic explanation.

The reduction of a psychological theory to a program for a Turing machine is a way of exorcising the homunculi. The reduction ensures that no operations have been postulated except those that could be performed by a familiar mechanism. Of course, the working psychologist usually cannot specify the reduction for each functionally individuated process in every theory he is prepared to take seriously. In practice the argument usually goes in the opposite direction; if the postulation of a mental operation is essential to some cherished

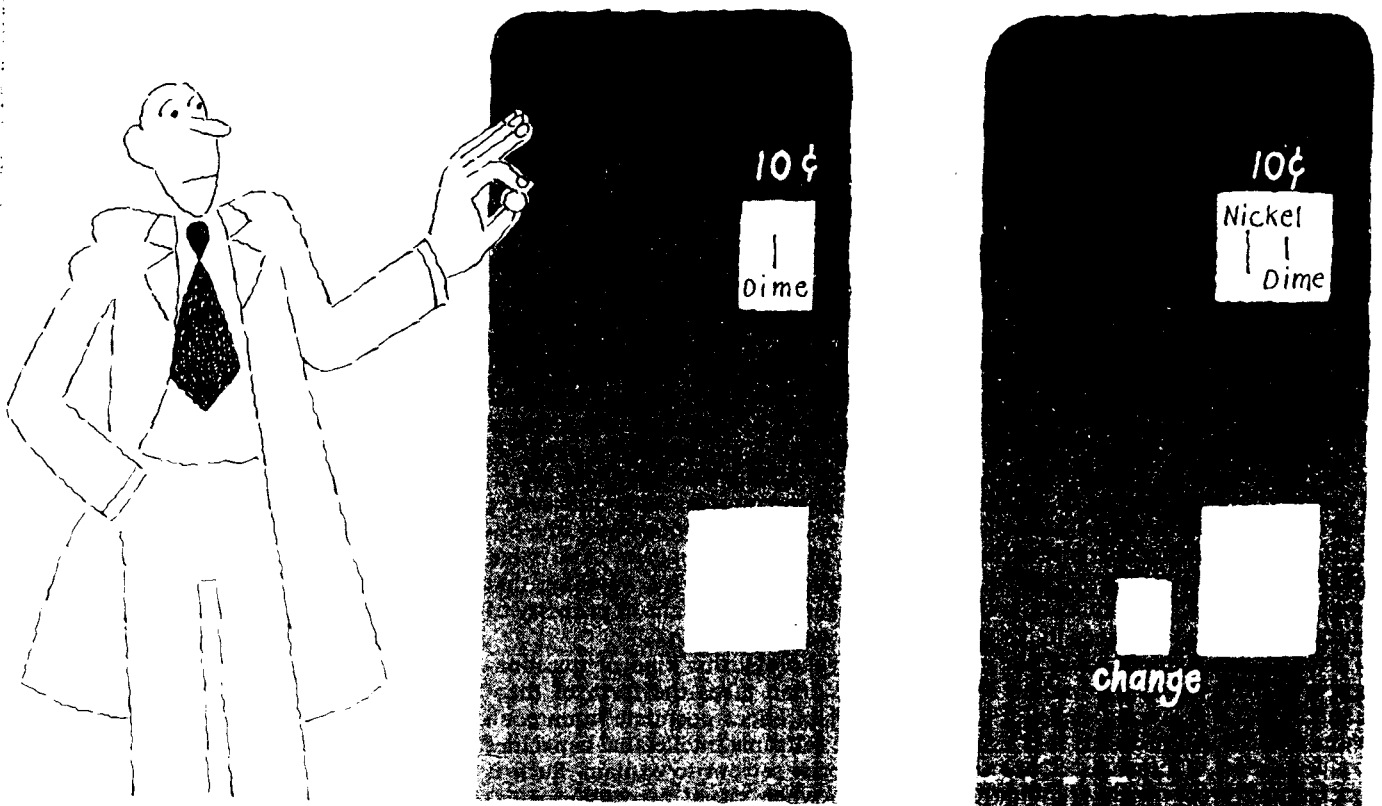
psychological explanation, the theorist tends to assume that there must be a program for a Turing machine that will carry out that operation.

The "black boxes" that are common in flow charts drawn by psychologists often serve to indicate postulated mental processes for which Turing reductions are wanting. Even so, the possibility in principle of such reductions serves as a methodological constraint on psychological theorizing by determining what functional definitions are to be allowed and what it would be like to know that everything has been explained that could possibly need explanation.

Such is the origin, the provenance and the promise of contemporary functionalism. How much has it actually paid off? This question is not easy to answer because much of what is now happening in the philosophy of mind and the cognitive sciences is directed at exploring the

	STATE S0
DIME INPUT	DISPENSES A COKE

	STATE S1	STATE S2
NICKEL INPUT	GIVES NO OUTPUT AND GOES TO S2	DISPENSES A COKE AND GOES TO S1
DIME INPUT	DISPENSES A COKE AND STAYS IN S1	DISPENSES A COKE AND A NICKEL AND GOES TO S1



TWO COKE MACHINES bring out the difference between behaviorism (the doctrine that there are no mental causes) and mentalism (the doctrine that there are mental causes). Both machines dispense a Coke for 10 cents and have states that are defined by reference to their causal role. The machine at the left is a behavioristic one: its single state (*S0*) is defined solely in terms of the input and the output.

The machine at the right is a mentalistic one: its two states (*S1*, *S2*) must be defined not only in terms of the input and the output but also in terms of each other. To put it another way, the output of the Coke machine depends on the state the machine is in as well as on the input. The functionalist philosopher maintains that mental states are interdefined, like the internal states of the mentalistic Coke machine.

scope and limits of the functionalist explanations of behavior. I shall, however, give a brief overview.

An obvious objection to functionalism as a theory of the mind is that the functionalist definition is not limited to mental states and processes. Catalysts, Coke machines, valve openers, pencil sharpeners, mousetraps and ministers of finance are all in one way or another concepts that are functionally defined, but none is a mental concept such as pain, belief and desire. What, then, characterizes the mental? And can it be captured in a functionalist framework?

The traditional view in the philosophy of mind has it that mental states are distinguished by their having what are called either qualitative content or intentional content. I shall discuss qualitative content first.

It is not easy to say what qualitative content is; indeed, according to some theories, it is not even possible to say what it is because it can be known not by description but only by direct experience. I shall nonetheless attempt to describe it. Try to imagine looking at a blank wall through a red filter. Now change the filter to a green one and leave everything else exactly the way it was. Something about the character of your experience changes when the filter does, and it is this kind of thing that philosophers call qualitative content. I am not entirely comfortable about introducing qualitative content in this way, but it is a subject with which many philosophers are not comfortable.

The reason qualitative content is a problem for functionalism is straightforward. Functionalism is committed to defining mental states in terms of their causes and effects. It seems, however, as if two mental states could have all the same causal relations and yet could differ in their qualitative content. Let me illustrate this with the classic puzzle of the inverted spectrum.

It seems possible to imagine two observers who are alike in all relevant psychological respects except that experiences having the qualitative content of red for one observer would have the qualitative content of green for the other. Nothing about their behavior need reveal the difference because both of them see ripe tomatoes and flaming sunsets as being similar in color and both of them call that color "red." Moreover, the causal connection between their (qualitatively distinct) experiences and their other mental states could also be identical. Perhaps they both think of Little Red Riding Hood when they see ripe tomatoes, feel depressed when they see the color green and so on. It seems as if anything that could be packed into the notion of the causal role of their experiences could be shared by them, and yet the qualitative content of the experi-

ences could be as different as you like. If this is possible, then the functionalist account does not work for mental states that have qualitative content. If one person is having a green experience while another person is having a red one, then surely they must be in different mental states.

The example of the inverted spectrum is more than a verbal puzzle. Having qualitative content is supposed to be a chief factor in what makes a mental state conscious. Many psychologists who are inclined to accept the functionalist framework are nonetheless worried about the failure of functionalism to reveal much about the nature of consciousness. Functionalists have made a few ingenious attempts to talk themselves and their colleagues out of this worry, but they have not, in my view, done so with much success. (For example, perhaps one is wrong in thinking one can imagine what an inverted spectrum would be like.) As matters stand, the problem of qualitative content poses a serious threat to the assertion that functionalism can provide a general theory of the mental.

Functionalism has fared much better with the intentional content of mental states. Indeed, it is here that the major achievements of recent cognitive science are found. To say that a mental state has intentional content is to say that it has certain semantic properties. For example, for Enrico to believe Galileo was Italian apparently involves a three-way relation between Enrico, a belief and a proposition that is the content of the belief (namely the proposition that Galileo was Italian). In particular it is an essential property of Enrico's belief that it is about Galileo (and not about, say, Newton) and that it is true if, and only if, Galileo was indeed Italian. Philosophers are divided on how these considerations fit together, but it is widely agreed that beliefs involve semantic properties such as expressing a proposition, being true or false and being about one thing rather than another.

It is important to understand the semantic properties of beliefs because theories in the cognitive sciences are largely about the beliefs organisms have. Theories of learning and perception, for example, are chiefly accounts of how the host of beliefs an organism has are determined by the character of its experiences and its genetic endowment. The functionalist account of mental states does not by itself provide the required insights. Mousetraps are functionally defined, yet mousetraps do not express propositions and they are not true or false.

There is at least one kind of thing other than a mental state that has intentional content: a symbol. Like thoughts, symbols seem to be about things. If

someone says "Galileo was Italian," his utterance, like Enrico's belief, expresses a proposition about Galileo that is true or false depending on Galileo's homeland. This parallel between the symbolic and the mental underlies the traditional quest for a unified treatment of language and mind. Cognitive science is now trying to provide such a treatment.

The basic concept is simple but striking. Assume that there are such things as mental symbols (mental representations) and that mental symbols have semantic properties. On this view having a belief involves being related to a mental symbol, and the belief inherits its semantic properties from the mental symbol that figures in the relation. Mental processes (thinking, perceiving, learning and so on) involve causal interactions among relational states such as having a belief. The semantic properties of the words and sentences we utter are in turn inherited from the semantic properties of the mental states that language expresses.

Associating the semantic properties of mental states with those of mental symbols is fully compatible with the computer metaphor, because it is natural to think of the computer as a mechanism that manipulates symbols. A computation is a causal chain of computer states and the links in the chain are operations on semantically interpreted formulas in a machine code. To think of a system (such as the nervous system) as a computer is to raise questions about the nature of the code in which it computes and the semantic properties of the symbols in the code. In fact, the analogy between minds and computers actually implies the postulation of mental symbols. There is no computation without representation.

The representational account of the mind, however, predates considerably the invention of the computing machine. It is a throwback to classical epistemology, which is a tradition that includes philosophers as diverse as John Locke, David Hume, George Berkeley, René Descartes, Immanuel Kant, John Stuart Mill and William James.

Hume, for one, developed a representational theory of the mind that included five points. First, there exist "Ideas," which are a species of mental symbol. Second, having a belief involves entertaining an Idea. Third, mental processes are causal associations of Ideas. Fourth, Ideas are like pictures. And fifth, Ideas have their semantic properties by virtue of what they resemble: the Idea of John is about John because it looks like him.

Contemporary cognitive psychologists do not accept the details of Hume's theory, although they endorse much of its spirit. Theories of computation provide a far richer account of mental processes than the mere association of Ideas. And only a few psychologists still think

that imagery is the chief vehicle of mental representation. Nevertheless, the most significant break with Hume's theory lies in the abandoning of resemblance as an explanation of the semantic properties of mental representations.

Many philosophers, starting with Berkeley, have argued that there is something seriously wrong with the suggestion that the semantic relation between a thought and what the thought is about could be one of resemblance. Consider the thought that John is tall. Clearly the thought is true only of the state of affairs consisting of John's being tall. A theory of the semantic properties of a thought should therefore explain how this particular thought is related to this particular state of affairs. According to the resemblance theory, entertaining the thought involves having a mental image that shows John to be tall. To put it another way, the relation between the thought that John is tall and his being tall is like the relation between a tall man and his portrait.

The difficulty with the resemblance theory is that any portrait showing John to be tall must also show him to be many other things: clothed or naked, lying, standing or sitting, having a head or not having one, and so on. A portrait of a tall man who is sitting down resembles a man's being seated as much as it resembles a man's being tall. On the resemblance theory it is not clear what distinguishes thoughts about John's height from thoughts about his posture.

The resemblance theory turns out to encounter paradoxes at every turn. The possibility of construing beliefs as involving relations to semantically interpreted mental representations clearly depends on having an acceptable account of where the semantic properties of the mental representations come from. If resemblance will not provide this account, what will?

The current idea is that the semantic properties of a mental representation are determined by aspects of its functional role. In other words, a sufficient condition for having semantic properties can be specified in causal terms. This is the connection between functionalism and the representational theory of the mind. Modern cognitive psychology rests largely on the hope that these two doctrines can be made to support each other.

No philosopher is now prepared to say exactly how the functional role of a mental representation determines its semantic properties. Nevertheless, the functionalist recognizes three types of causal relation among psychological states involving mental representations, and they might serve to fix the semantic properties of mental representations. The three types are causal relations among mental states and stimuli, mental

states and responses and some mental states and other ones.

Consider the belief that John is tall. Presumably the following facts, which correspond respectively to the three types of causal relation, are relevant to determining the semantic properties of the mental representation involved in the belief. First, the belief is a normal effect of certain stimulations, such as seeing John in circumstances that reveal his height. Second, the belief is the normal cause of certain behavioral effects, such as uttering "John is tall." Third, the belief is a normal cause of certain other beliefs and a normal effect of certain other beliefs. For example, anyone who believes John is tall is very likely also to believe someone is tall. Having the first belief is normally causally sufficient for having the second belief. And anyone who believes everyone in the room is tall and also believes John is in the room will very likely believe John is tall. The third belief is a normal effect of the first two. In short, the functionalist maintains that the proposition expressed by a given mental representation depends on the causal properties of the mental states in which that mental representation figures.

The concept that the semantic properties of mental representations are determined by aspects of their functional role is at the center of current work in the cognitive sciences. Nevertheless, the concept may not be true. Many philosophers who are unsympathetic to the cognitive turn in modern psychology doubt its truth, and many psychologists would probably reject it in the bald and unelaborated way that I have sketched it. Yet even in its skeletal form, there is this much to be said in its favor: It legitimizes the notion of mental representation, which has become increasingly important to theorizing in every branch of the cognitive sciences. Recent advances in formulating and testing hypotheses about the character of mental representations in fields ranging from phonetics to computer vision suggest that the concept of mental representation is fundamental to empirical theories of the mind.

The behaviorist has rejected the appeal to mental representation because it runs counter to his view of the explanatory mechanisms that can figure in psychological theories. Nevertheless, the science of mental representation is now flourishing. The history of science reveals that when a successful theory comes into conflict with a methodological scruple, it is generally the scruple that gives way. Accordingly the functionalist has relaxed the behaviorist constraints on psychological explanations. There is probably no better way to decide what is methodologically permissible in science than by investigating what successful science requires.



A little girl shouldn't have to beg for food.

But Nita must.

Her frail mother, who spends all day in the marketplace peddling straw mats, can't sell enough to feed Nita and her two younger brothers.

For \$15 a month through our sponsorship program, you can help a child like Nita. For a destitute child, your generosity can mean health, an education—even life itself.

Write to: Mrs. Jeanne Clarke Wood,
Children, Incorporated, P.O. Box 5381,
Dept. SA1Y, Richmond, Va. 23220 USA

- I wish to "adopt" a boy girl in Asia, Latin America, Middle East, Africa, USA, Greatest Need.
- I will give \$15 a month (\$180 a year). Enclosed is my gift for a full year the first month . Please send me the child's name, story, address and picture.
- I can't "adopt," but will help \$ _____
- Please send me further information.
- If for a group, please specify _____

Church, Class, Club, School, Business, etc.

NAME _____

ADDRESS _____

CITY _____

STATE _____

ZIP _____

U.S. gifts are fully tax deductible.
Annual financial statements are available on request.

CHILDREN, INC.