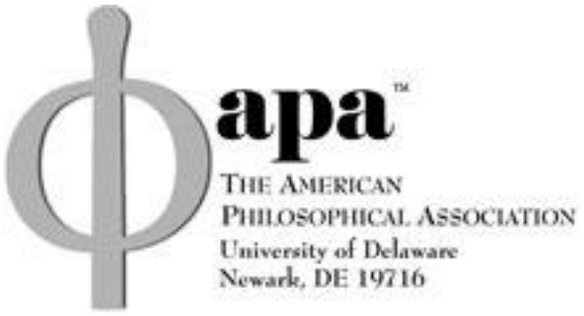

Machines and the Mental

Author(s): Fred Dretske

Source: *Proceedings and Addresses of the American Philosophical Association*, Sep., 1985, Vol. 59, No. 1 (Sep., 1985), pp. 23-33

Published by: American Philosophical Association

Stable URL: https://www.jstor.org/stable/3131645

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/3131645?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

American Philosophical Association is collaborating with JSTOR to digitize, preserve and extend access to *Proceedings and Addresses of the American Philosophical Association*

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

MACHINES AND THE MENTAL[1]*

Fred Dretske
University of Wisconsin/Madison

Computers are machines and there are a lot of things machines can't do. But there are a lot of things *I* can't do: speak Turkish, understand James Joyce, or recognize a nasturtium when I see one. Yet, numerous as are my disabilities, they do not materially affect my status as a thinking being. I lack specialized skills, knowledge and understanding, but nothing that is essential to membership in the society of rational agents. With machines, though, and this includes the most sophisticated modern computers, it is different. They *do* lack something that is essential.

Or so some say. And so say I. In saying it, though, one should, as a philosopher, be prepared to say what *is* essential, what are the conditions for membership in this exclusive club. If an ability to understand James Joyce isn't required, what, then, *must* one be able to understand? If one doesn't have to know what nasturtiums look like, is there something else one must be able to identify? What might this be? If one is told that there is no specific thing one has to understand, identify or know but, nonetheless, something *or other* towards which one must have a degree of competence, it is hard to see how to deny computers admission to the club. For even the simple robots designed for home amusement talk, see, remember and learn. Or so I keep reading in the promotional catalogs. Isn't this enough? Why not?

I happen to be one of those philosophers who, though happy to admit that minds computer, and in this sense *are* computers, have great difficulty seeing how computers could be minded. I'm not (not *now* at least) going to complain about the impoverished inner life of the computer--how they don't feel pain, fear, love or anger. Nor am I going to talk about the mysterious inner light of consciousness. For I'm not at all sure one needs feelings or self consciousness to solve problems, play games, recognize patterns and understand stories. Why can't pure thought, the sort of thing computers purportedly have, stand to ordinary thought, the sort of thing we have, the way a solitary stroll stands to a hectic walk down a crowded street? The same thing--walking--is going on in both cases. It just *seems* different because, in the latter case, so much else is going on at the same time. A mathematician's calculations are no less brilliant, certainly no less deserving of classification as mental, because he or she is blind, deaf and emotionally stunted--because, in other words,

---

*Presidential Address delivered before the Eighty-third Annual Meeting of the Western Division of the American Philosophical Association, Chicago, Illinois, April 26, 1985.

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

the calculations occur within a comparatively anemic sensory and emotional environment. Why can't we think of our machines as occupying a position on the far right of this mental continuum? Just a bit to the right of Star Trek's Doctor Spock? We don't, after all, deny someone the capacity for love because they can't do differential calculus. Why deny the computer the ability to solve problems or understand stories because it doesn't feel love, experience nausea, or suffer indigestion?

Nor am I going to talk about how bad computers are at doing what most children can do--e.g., speak and understand their native language, make up a story or appreciate a joke. For such comparisons make it sound like a competition, a competition in which humans, with their enormous head start, and barring dramatic breakthroughs in AI, will remain unchallenged for the foreseeable future. I don't think the comparison should be put in these terms because I don't think there is a genuine competition in this area at all. It isn't that the best machines are still at the level of two-year-olds, requiring only greater storage capacity and fancier programming to grow up. Nor should we think of them as idiot savants, exhibiting a spectacular ability in a few isolated areas, but having an overall IQ too low for fraternal association. For machines, even the best of them, don't have an IQ. They don't *do* what we do--at least none of the things that, when we do them, exhibit intelligence. And it's not just that they don't do them the way we do them or as well as we do them. They don't do them at all. They don't solve problems, play games, prove theorems, recognize patterns, let alone think, see and remember. They don't even add and subtract.

To convince you of this, it is useful to look at our relationship to various instruments and tools. The preliminary examination will not take us far, but it will set the stage for a clearer statement of what I take to be the fundamental difference between minds and machines.

In our descriptions of instruments and tools we tend to assign them the capacities and powers of the agents who use them. We often think, or at least talk, of artifacts--tools, instruments and machines--as telling us things, recognizing, sensing, remembering and, in general, doing things that, in our more serious, literal, moments, we acknowledge to be the province of rational agents. In most cases this figurative use of language does no harm. No one is really confused. Though we open doors, and keys open (locked) doors, no one seems to worry about whether keys open doors better than we do, whether we are still ahead in this competition. No one is trying to build a fifth generation key that will surpass us in this enterprise. Why not? Since both keys and people open doors, why doesn't it make sense to ask who does it better? Because, of course, we all understand that doors are opened *with* keys. *We* are the agents. The key is the instrument. That we sometimes speak of the instrument in terms appropriate to the agent, speak of the key as doing what the agent does with the key, should not tempt us into supposing that, therefore, there are some things we do that keys can also do. We catch fish *with* worms; *we*, not the worms, catch the fish.

Before concluding, however, that computers are, like keys, merely fancy instruments in our cognitive tool box--and, thus, taken by themselves, unable to do what we can do with them--consider another case. Who really picks up the dust, the maid or the vacuum cleaner? Is the vacuum cleaner merely an instrument that the maid uses to pick up dust? Well yes, but not *quite* the way one uses a key to open a door or a hammer to pound a nail. One pushes the vacuum cleaner around but *it* picks

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

up the dust. In this case (unlike the key case) the question: "Who picks up dust better: people or vacuum cleaners?" *does* make good sense, and the answer, obviously, is the vacuum cleaner. We may never have had any real competition from keys for opening doors, but we seem to have lost the race for picking up dust to vacuum cleaners.

What such examples reveal is that the agent-instrument distinction is no certain guide to who or what is to be given credit for a performance. We do things. Machines do things. Sometimes we do things with machines. Who gets the credit depends on what is done and how it is done. To ask whether a simple pocket calculator can really multiply or whether it is *we* who multiply *with* the calculator is to ask, whether, relative to this task, the agent-instrument relation is more like our use of a key in opening a door or more like our use of a vacuum cleaner in picking up dust.

Well, then, are computers our computational keys? Or are they more like vacuum cleaners? Do they literally do the computational tasks that we sometimes do without them but do it better, faster, and more reliably? This may sound like a rather simple-minded way to approach the issue of minds and machines, but unless one gets clear about the relatively simple question of *who* does the job, the person or the pocket calculator, in adding up a column of figures, one is unlikely to make such progress in penetrating the more baffling question of whether more sophisticated machines exhibit (or will some day) some of the genuine qualities of intelligence. For I assume that if machines can really play chess, prove theorems, understand a text, diagnose an illness, and recognize an object--all achievements that are routinely credited to modern machines by sober members of the artificial intelligence community--if these descriptions are *literally* true, then to that degree they participate in the intellectual enterprise. To that degree they are minded. To that extent they belong in the club however much we, with our prejudice in favor of biological look-alikes, may continue to deny them full admission.

So let me begin with a naive question: Can computers add? We may not feel very threatened if this is *all* they can do. Nevertheless, if they do even this much, then the barriers separating mind and machine have been breached and there is no reason to think they won't eventually be removed.

The following argument is an attempt to show that whatever it is that computers are doing when we use them to answer our arithmetical questions, it isn't addition. Addition is an operation on numbers. We add 7 and 5 to get 12, and 7, 5 and 12 are numbers. The operations computers perform, however, are not operations on numbers. At best, they are operations on certain physical tokens that *stand for*, or are interpreted as standing for, the numbers. Therefore, computers don't add.

In thinking about this argument (longer than I care to admit) I decided that there was something right about it. *And* something wrong. What is right about it is the perfectly valid (and relevant) distinction it invokes between a representation and what it represents, between a sign and what it signifies, between a symbol and its meaning or reference. We have various ways of representing or designating the numbers. The written numeral "2" stands for the number 2. So does "two." Unless equipped with special pattern recognition capabilities, machines are not prepared to handle these particular symbols (the symbols appear on the keyboard for *our* convenience). But they have their own system of representation: open and closed switches, the orientation of magnetic fields, the distribution of holes on a card. But

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

whatever the form of representation, the machine is obviously restricted to operations on the symbols or representations themselves. It has no access, so to speak, to the *meaning* of these symbols, to the things the representations represent, to the numbers. When instructed to add two numbers stored in memory, the machine manipulates representations in some electromechanical way until it arrives at another representation--something that (if things go right) stands for the sum of what the first two representations stood for. At no point in the proceedings do numbers, in contrast to numerals, get involved. And if, in order to add two numbers, one has to perform some operation on the numbers themselves, then what the computer is doing is not addition at all.

This argument, as I am sure everyone is aware, shows *too* much. It shows that *we* don't add either. For whatever operations may be performed in or by our central nervous system when we add two numbers, it quite clearly isn't an operation on the numbers themselves. Brains have their own coding systems, their own way of representing the objects (including the numbers) about which its (or our) thoughts and calculations are directed. In this respect a person is no different than a computer. Biological systems may have different ways of representing the objects of thought, but they, like the computer, are necessarily limited to manipulating these representations. This is merely to acknowledge the nature of thought itself. It is a *vicarious* business, a *symbolic* activity. Adding two numbers is a way of thinking *about* two numbers, and thinking *about* X and Y is not a way of pushing X and Y around. It is a way of pushing around their symbolic representatives.

What is wrong with the argument, then, is the assumption that in order to add two numbers, a system must literally perform some operation on the numbers themselves. What the argument shows, if it shows anything, is that in order to carry out arithmetical operations, a system must have a way of representing the numbers and the capacity for manipulating these representations in accordance with arithmetic principles. But isn't this precisely what computers have?

I have discussed this argument at some length only to make the point that all cognitive operations (whether by artifacts or natural biological systems) will necessarily be realized in some electrical, chemical or mechanical operation over physical structures. (Or, if materialism isn't true, they will be realized in or by transformations of mind-stuff.) This fact alone doesn't tell us anything about the cognitive nature of the operations being performed--whether, for instance, it is an inference, a thought or the taking of a square root. For what makes these operations into thoughts, inferences, or arithmetical calculations is, among other things, the meaning or, if you prefer, the semantics of those structures over which they are performed. To think about the number 7 or your cousin George, you needn't do anything with the number 7 or your cousin George, but you do need the internal resources for representing 7 and George and the capacity for manipulating these representations in ways that stand for activities and conditions of the things being represented.

This should be obvious enough. Opening and closing relays doesn't count as addition, or as moves in a chess game, unless the relays, or their various states, stand for numbers and chess moves. But what may not be so obvious is that these physical activities cannot acquire the relevant kind of meaning merely by *assigning* them an interpretation, by letting them mean something *to* or *for us.* Unless the symbols being manipulated mean something *to the system manipulating them,* their meaning,

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

whatever it is, is irrelevant to evaluating what the system is doing when it manipulates them.[2] I cannot make you, someone's parrot, or a machine think about my cousin George, or the number 7, just by assigning meanings in accordance with which this is what your (the parrot's, the machine's) activities stand for. If things were this easy, I could make a tape recorder think about my cousin George. Everything depends on whether this is the meaning these events have to you, the parrot, or the machine.

Despite some people's tendency to think that the manipulation of symbols is *itself* a wondrous feat, worthy of such inflated descriptions as "adding numbers," "drawing conclusions," or "figuring out its next move" the process is, in fact, absolutely devoid of cognitive significance.[3] I once watched a gerbil manipulate a symbol, a symbol that, according to conventional standards--standards that I, but not the gerbil--understood, stood for my bank balance. I didn't have the slightest temptation to see in this symbol manipulation (actually *consumption*) process anything of special significance. Even if I trained a fleet of gerbils to arrange symbols in some computationally satisfying way (e.g., to balance my checkbook), I don't think *they* should be credited with balancing my checkbook. *I* would merely be using the gerbils to balance my checkbook in the way I use worms to catch fish.

To understand *what* a system is doing when it manipulates symbols, it is necessary to know, not just what these symbols mean, what interpretation they have been, or can be, *assigned*, but what they mean to the system performing the operations. John Searle and Ned Block have dramatized this point.[4] Searle, for instance, asks one to imagine someone who understands no Chinese manipulating Chinese symbols in accordance with rules expressed in a language he does understand. Imagine the rules cleverly enough designed so that this person can carry on a correspondence in Chinese--responding to (written) Chinese questions with (written) Chinese answers in a way that is indistinguishable from the performance of a native speaker of Chinese. Clearly, though a correspondent might not be able to discover this fact, the symbol manipulator himself doesn't understand Chinese. Nor does the system of which he is a part. Understanding Chinese is not just a matter of manipulating meaningful symbols in some appropriate way. These symbols must mean something *to* the system performing the operations.

This should not be taken to imply that machines cannot serve as useful models for cognitive processes. On the contrary. Their prevalent use in cognitive psychology indicates otherwise. What it does imply is that the machines do not literally *do* what we do when we engage in those activities for which they provide an effective model. Computer simulations of a hurricane do not blow trees down. Why should anyone suppose that computer simulations of problem solving must themselves solve problems?

But how does one build a system that is capable not only of performing operations on (or with) symbols, but one *to which* these symbols mean something, a machine that, in this sense, understands the meaning of the symbols it manipulates? Only when we can do this will we have machines that not only produce meaningful output, but machines whose activities in producing that output bear the mark of the mental. Only then will we have machines that we can not only *use* to balance our checkbook, but machines that will do it for us, machines that will not only print out answers to our questions, but machines that will *answer* our questions.

One thing seems reasonably clear: if the meaning of the symbols on which a machine performs its operations is a meaning wholly derived from us, its users--if

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

it is a meaning that *we* assign the various states of the machine and, therefore, a meaning that we can change at will without altering the *way* these symbols are processed by the machine itself--then there is no way the machine can acquire understanding, no way these symbols can have a meaning to *the machine itself.* Unless these symbols have what we might call an intrinsic meaning, a meaning they possess which is independent of our communicative intentions and purposes, then this meaning *must* be irrelevant to assessing what the machine is doing when it manipulates them. The machine is processing meaningful (to us) symbols, to be sure, but the *way* it processes them is quite independent of *what they mean*--hence, nothing *the machine* does is explicable in terms of the meaning of the symbols it manipulates or, indeed, of their even having a meaning. Given the right programming and data base, we can contrive to make the sentences a machine produces answers to our questions. But the machine itself is no more answering our questions than is an automatic teller (now so prevalent in the banking industry) embezzling money when it keeps our deposit without crediting our account.

In order, therefore, to approximate something of genuine cognitive significance, in order to give a machine something that bears *a* mark, if not *all* the marks, of the mental, the symbols a machine manipulates must be given a meaning of their own, a meaning that is independent of their user's purposes and intentions. Only by doing this will it become possible to make the meaning of these symbols relevant to what the machine does with them, possible, in other words, to make the machine do something *because* of what its symbols mean, possible, therefore, to make these symbols mean something to the machine itself.

And how might this be done? In the same way, I submit, that nature arranged it in our case. We must put the computer into the head of a robot, into a larger system that has the kind of sensory capabilities, the perceptual resources, that enable what goes on inside the computer to mean something, in Paul Grice's natural sense of meaning[5], about what goes on outside the computer. The elements over which the computer performs its operations will then have a meaning that is independent of the conventions of its users. They will then mean something in the same way the swing of a galvanometer needle means something regarding the electrical activity in the circuit to which it is connected, the way expanding mercury means something about the surrounding temperature, the way a voltage spike in our visual cortex means something about the distribution of light impinging on the retina. This kind of meaning is sometimes called information.[6] It is the kind of meaning we associate with reliable signs and trustworthy indicators, the kind of meaning possessed by dark clouds, shadows, prints, leaf patterns, smoke, acoustic vibrations, and the electrical activity in the sensory pathways. The difference between a robot and the disembodied computer found in our office buildings and laboratories is that the former, unlike the latter, have symbol systems that are also *sign* systems: signs being symbols having a meaning quite independent of what *we* might say or think they mean. The only intrinsic meaning in most computers is the meaning derived from the array of pressure sensitive transducers on its keyboard. The activities in the computer may mean a move to KB-3 *to us*, but all they mean *to the computer* is that key 37 has been depressed.

This is only to say that information, *real* information, the kind of meaning associated with natural signs, is irrelevant to the operation of high speed digital computers in a way it is not irrelevant to the operation of living systems. If a sea snail doesn't

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

get information about the turbulence in the water, if there isn't some state *in* the snail that functions as a natural sign of turbulent water, it risks being dashed to pieces when it swims to the surface to obtain the micro-organisms on which it feeds. If (certain) bacteria did not have something inside that meant that *that* was the direction of magnetic north, they could not orient themselves so as to avoid toxic surface water. They would perish. If, in other words, an animal's internal sensory states were not rich in information, intrinsic natural meaning, about the presence of prey, predators, cliffs, obstacles, water and heat, it could not survive. It isn't enough to have the internal states of these creatures mean something *to us*, for it to have *symbols* it can manipulate. If these symbols don't somehow register the conditions in their possessor's surroundings, the creature's symbol manipulation capacity is completely worthless. Of what possible significance is it to be able to handle symbols for food, danger and sexual mates if the occurrence of these symbols is wholly unrelated to the actual presence of food, danger and sexual mates?

In a sense, then, work on machine perception, pattern recognition, and robotics has greater relevance to the cognitive capacities of machines than the most sophisticated programming in such purely intellectual tasks as language translation, theorem proving, or game playing. For a pattern recognition device is at least a device whose internal states, like those of the bacterium, snail and human being, mean something about what is happening, or the conditions that exist, around it. There is actually something *in* these machines that means something regarding what is happening outside them and, moreover, something that means this whether or not we, the users of the machine (or, indeed, the machine itself), recognize it. We are not free to assign or withhold this meaning--anymore than we are free to say what the screech of a smoke alarm means. We can *say* that the alarm means there are leopards nearby, and for certain purposes (e.g., in a children's game of make-believe) we may even want to give it that meaning. But that isn't actually what the sound means. That isn't what it is a sign of, not the information it carries. And for the same reason, the meaning of the internal states of a pattern recognition device, or a robot equipped with sensory capacities, is a meaning these states have which, if it isn't actually a meaning *for* the machine itself, is the only meaning that shows any promise of being promoted into something that is relevant to assessing what these machines are doing when they mobilize these meaningful elements to produce an output.

But have we come any closer to understanding genuine mentation, the capacity to add, subtract, plan, play games, understand stories, and think about one's cousin George? What we have so far required of any aspiring symbol-manipulator is, in effect, that *some* of its symbols be actual signs of the conditions they signify, that there *be* some system-to-world correlations that confer on these symbols an intrinsic meaning, a meaning they do not derive wholly from the purposes and intentions of their users. This puts the symbol-manipulator *in the world* in a way it would not otherwise be. But have we come any closer to understanding how an element, symbol *or* sign, could have meaning *to* the symbol manipulator itself, how this meaning, and not just the sign having this meaning, could be relevant to *what* the system is doing when it manipulates these signs?

Think about a dog that has been trained to detect marijuana. Custom's agents can use these dogs to find concealed marijuana. When the dog barks, wags its tail, or does whatever it was trained to do when it smells marijuana, this alerts the agent to its presence. As a result of the dog's behavior, the official comes to believe that

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

there is marijuana in the suitcase. But what does *the dog* believe? Surely not what the agent believes--*viz*., that there is marijuana in the suitcase. Why not? There is obviously something in the dog that is sensitive to the presence of marijuana, some neural condition whose occurrence is a sign, and in this sense means, that there is marijuana nearby. Furthermore, this something is (as a result of training) getting the dog to wag its tail or bark. Why isn't this enough to justify attributing a belief to the dog, a belief with the content: there is marijuana nearby? If we had a Stanford robot that could perform half as well with blocks on a table, we would doubtless be hearing about its extraordinary recognitional capacities. But nobody seems terribly impressed with the dog. The dog, one can hear them saying, has a wonderfully discriminating sense of smell. It has *sensory* powers that exceed those of its trainers, but its *conceptual* or *cognitive* capacities are modest indeed. It can smell marijuana, sure enough. It can even be trained to respond in some distinctive way to this smell. But it doesn't have the conceptual resources for believing *of* what it smells *that* it is marijuana.

If we are going to treat the dog in this deflationary way, we should be prepared to do the same with machines--including fancy robots. In industrial applications of machine vision, for example, it is said that machines can recognize short circuits on the printed circuit boards they examine. Not so. The machine merely searches for breaks or discontinuities in the metallic deposit. It is concerned with *spatial* discontinuities. We, its users, are worried about electrical discontinuities. Under the right circumstances, we can use something that detects the first as an instrument, a means, for identifying the second, but, just like the dog, the instrument should not be credited with the *conceptual* talents of its users, what *we* are able to discover by using it. The machine is no more able to have electricity thoughts than the dog is able to have marijuana thoughts.

Some people think that what machines lack is conscious awareness. Perhaps they do. But our marijuana sniffing dog should teach us that this isn't the missing ingredient, not what we need to manufacture a thinker of thoughts out of a sign manipulator. For the dog *is*, whereas the custom's agent is not, *aware of* the concealed marijuana. The dog smells it and the agent does not. To give a system the kind of meaning we now seek, to give it genuine understanding, it is not enough to give it conscious awareness *of* the stuff it is supposed to cognize. It isn't even enough to make the creature's conscious awareness of the stuff *cause* it to behave in some appropriate way *toward* the stuff. For, as our trained dog illustrates, all this can be true without the system, animal or machine, having the slightest conception of what it all means. And what we are after, of course, is something that wags its tail, activates its printer, or starts its motors, not just because it is aware of, say, marijuana, but because it thinks, judges, or believes that it is marijuana. What we are after is *con*ception, not *per*ception.

The difference between machines (or dogs) and the agents who use them is that although machines (and dogs) can pick up, process and transmit the information we need in our investigative efforts (this is what makes them useful tools), although they can respond (either by training or programming) to meaningful signs, it isn't the meaning of the signs that figures in the explanation of why they do what they do. Some internal sign of marijuana, some neurological condition that, in this sense, means that marijuana is present, can cause the dog's tail to move, but it isn't the fact that it means this that explains the tail movement. This, I submit, is the differ-

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

ence between the dog and its master, between the machine and its users, between the robot and the people it replaces. When I smell marijuana, my finger wagging is produced, not simply (as in the case of the dog) by a neurological condition that means that marijuana is present, but by the *meaning* of this neurological condition, by the fact that it means this and not something else. In *my* case the motor activity is produced by the meaning of an occurrent sign; in the dog's case by the occurrence of a sign having that meaning. To say that the smell of marijuana means something to me that it doesn't mean to the dog is merely to say that its meaning what it does makes a difference to what I do but not to what the dog does. That is why it is true of me, but not the dog, that I wag my finger because I think marijuana is present, because I am in an internal state having this content. The dog is in a state with the same content, to be sure, but it isn't this content that wags the tail. The difference between a thinker of marijuana thoughts (me) and the mere detector of marijuana (dog or machine) is not, then, merely a difference in what our internal signs mean, but a difference in whether, and if so, how, these meanings are implicated in the management of the signs themselves.

I seem to have painted myself into a corner. At least I expect to be told as much by those philosophers who are deeply suspicious of meaning. I expect to be told that meaning is an abstraction, not something that *could* play a role in the activities of a symbol manipulating system. From the control point of view, meaning is an epiphenomenon. It is causally inert. Even if one agrees that there are signs in the head, it is the signs themselves, not their meaning, that turn the cranks, pull the levers, and depress the accelerator. It is the grey stuff inside, not what it means, that activates the motor neurons. Just ask the neurobiologists. If, in order to promote a processor of meaningful signs into a system with genuine understanding, into a real thinker of thoughts, we must give the meaning of these signs a role to play in the *way* these signs are processed, in the way the motor control system operates with them, then the prospects for effecting such a promotion, not just for machines, but for human beings as well, look bleak indeed.

Such pessimism, though widespread these days, is unwarranted. Meanings, of the kind now in question, *are* what philosophers like to call abstract entities, but they are *no more* abstract, and certainly no less capable of exercising a causal influence, than are, say, differences in weight, brightness and orientation. Just as the difference in weight between a basketball and a bowling ball may be responsible, causally responsible, for the behavior of a beam balance, the correlations constituting the meaning of a sign can, and regularly *do*, affect the way a system processes that sign. The correlation between a ringing bell and someone's presence at one's door, the kind of correlation that confers on the ringing bell the meaning that someone is at the door, *changes* the way a (suitably exposed) nervous system processes the internal sign of a ringing bell. Exposure (either directly or indirectly) to this correlation produces a difference in whether, and if so, which, motor neurons are activated by the internal sensory sign of a ringing bell. This, it seems to me, is a case where the meaning of a sign, and not just the sign that has that meaning, makes a difference in *how* a system processes that sign--hence, a case where the sign's meaning, and not the sign itself, helps to explain the behavior of the system in which that sign occurs.

My doorbell example is a homely example of the causal role of meaning. Some may think it ignores all the interesting questions. For it involves an agent already

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

possessed of the conceptual resources for interpreting signs, understanding meanings, and modifying his or her behavior in the light of experienced correlations. This is true, but irrelevant. For the very same phenomenon can be illustrated at almost every biological level, every level at which *learning* occurs. It is, in fact, merely an instance of what learning theorists describe as the contingencies modifying the way a system processes, and hence responds to, the internal signs for stimulus conditions. Even the lowly snail mentioned earlier changes the way it processes signs by exposing it to the correlations constituting the meaning of these signs. And it is, surely, the fact that our internal states are correlated with certain kinds of external conditions that helps to determine the ultimate outcome of the motor activities produced by these internal states. It is the correlations, therefore, that help to determine what kind of feedback we received from such activities and, hence, the likelihood of our repeating them in the same circumstances. It is, therefore, the correlations, not merely the internal correlates, that shape--hence, *explain*--learned behavior. Learning, in fact, *is* a process in which the meaning of internal signs (i.e., their correlation with external conditions), not (merely) the signs themselves, helps to determine how these signs are exploited for purposes of motor control. For such systems the internal signs not only have meaning, this meaning affects the way the system manages these signs; and it is in this sense that the signs mean something *to* the system in which they occur.

This, it seems to me, is a fundamental difference between the sign processing capabilities of various systems. It is a difference that helps explain why it seems so natural to say of some of them (human beings and some animals) but not others (machines and simple organisms) that the symbols they manipulate mean something to the symbol manipulator. It is a difference, I submit, that underlies our conviction that we, but not the machines and a variety of simple organisms, are genuine thinkers of thoughts. What *gives* us the capacities underlying this difference is a long and complicated story. It involves, I think, issues in learning theory, our multiple sensory access to the things we require to satisfy our needs, and the kind of feedback mechanisms we possess that allow us to modify *how* we manipulate internal signs by the kind of results our previous manipulations have produced. But this, clearly, is a story that we expect to hear from neurobiologists, not from philosophers. All I have been trying to tell is a simpler story, a story about the entrance requirements for admission to the club. I leave it to others to worry about how different systems manage, each in their own way, to satisfy these requirements.

Footnotes

1. My thanks to Denny Stampe for careful criticism and many useful suggestions. I also want to acknowledge the help given me by Fred Adams and the other sceptics in the audience at Augustana College where I read an early draft of this paper. They convinced me that the draft I read them was *earlier* than I ever suspected.

2. This is what Haugeland calls "original intentionality," something that, according to Haugeland, computers don't have: "To put it bluntly: computers themselves don't mean anything by their tokens (any more than books do)--they only mean what we say they do. Genuine understanding, on the other hand, is intentional

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms

in its own right" and not derivatively from something else. *Mind Design*, John Haugeland (ed.), Bradford Books; Montgomery, Vt., 1981, pp. 32-33. A number of authors have made essentially this point in their own way; e.g., Jerry Fodor, "Tom Swift and His Procedural Grandmother," *Cognition*, 6 (1978), reprinted in *Representations*, MIT/Bradford, 1981; Hilary Putnam, "Brains in a Vat," *Reason, Truth and History*, Cambridge University Press, 1981, pp. 10-11; Rob Cummins, *The Nature of Psychological Explanation*, MIT/Bradford, 1983, p. 94; Tyler Burge, "Belief *De Re*," *The Journal of Philosophy,* LXXILV, 6 (1977); John Searle, "Minds, Brains and Programs," *The Behavioral and Brain Sciences* 3:3 (1980).

3. In explaining why he thinks computers *can* (or will someday), Marvin Minsky (in "Why People Think Computers Can't," *AI Magazine*, Fall 1982), seems most impressed, for instance, with the fact that "computers can manipulate *symbols*."
4. John Searle, "Minds, Brains and Programs," *The Behavioral and Brain Sciences*, 3:3 (1980); Ned Block, "Troubles with Functionalism," in Wade Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minnesota Studies in the Philosophy of Science, Vol. 9, Minneapolis, Minn.; 1978.
5. Paul Grice, "Meaning," *Philosophical Review*, vol. 66 (1957), pp. 377-388.
6. See my *Knowledge and the Flow of Information,* MIT/Bradford: Cambridge, Mass., 1981.

This content downloaded from
132.174.254.159 on Tue, 24 Sep 2024 03:39:01 UTC
All use subject to https://about.jstor.org/terms