THE MECHANICAL MIND

Horace Barlow

Physiological Laboratory, Cambridge, CB2 3EG, England

Most neuroscientists accept the machine as a useful metaphor or model of
the mind. It points our research in a direction that has been outstandingly
successful for more than a century, namely the reductionist analysis of
brain function in terms of simpler physical, chemical, and biological pro-
cesses, and because we can understand all, or almost all, about machines,
the metaphor encourages us to think we can discover all, or almost all,
about the mind. I thought when I started writing this piece that the
metaphor was sound and useful, though its uncritical acceptance made me
uneasy, especially because I knew that many of my colleagues in subjects
like mathematics and linguistics received it with something close to incred-
ulity, while many others disliked it intensely. I therefore thought it would
be worthwhile to examine the idea in more detail, to try to find if there
was any justification for unease, incredulity, or dislike; if you read on you
will find that I have been forced to the conclusion that the metaphor is
misleading and potentially harmful, but that this results from prevalent
ignorance and prejudice about machines rather than any gross defect of
the metaphor.

IN DEFENCE OF THE ANALOGY

I think the main purpose behind calling the mind a machine is to drive out
demons. We are saying, in effect, "Look, there is no more to the working
of the brain than the physics and chemistry of its components, just as there
is no more to a machine than the physics and chemistry of its components."
This is admirable as an invocation not to waste time on mental spirits, and
to study the physics and chemistry instead, so what are the objections?
First there are three minor ones, namely that we don't actually know all
about machines, that brains are made of totally different materials, and
that they have come into existence in a strikingly different manner. By
considering these objections we see where to be cautious about the meta-
phor, but they are not fatal to it. However, there are more serious problems
that do justify mistrust and dislike: first, minds do some things that no
current machines do, and if one is mainly interested in these particular
tasks the analogy is not much help and could be misleading; second, it is
the minds of other people one interacts with, so to treat minds as machines
valued mainly for their usefulness has unpleasant ethical implications.

We Do Not Understand All About Machines

The person who understands most about a machine is its designer, and no
designer of a complex machine would claim that everything about it was
perfectly understood. Engineers are really very different from scientists, for
they use a body of knowledge to create something new, and provided the
goal is reached they are not too concerned if their creation has unforeseen
or unknown properties. Scientists make use of the same body of knowledge,
but they must pay particular attention to anything unforeseen or unknown,
since their goal is to extend the body of knowledge rather than to exploit
it. In engineering, intuitive leaps and creative solutions are rightly admired
when they make a machine work, but as scientists, we should admire
creativity and intuition that enable us to understand what was previously
unknown; we do not want to import the attitude "If it works, that's good
enough" along with the concept that mind is a product of engineering.

The current enthusiasm for neural networks may show that this niggle
is partly justified. Here are techniques that enable machines to perform
tasks that hitherto lay in the province of the mind, but whereas artificial
intelligence and the previous styles of computer simulation led to a more
detailed analytical knowledge of the requirements for performing the task,
the network style of simulation does not; instead it uses some blind pro-
cedure, such as back-propagation, and the success of a simulation is taken
as justification for the procedure, rather than for the programmer's analysis
of the task requirements. To be fair, the major proponents of the network
approach do not regard this as an advantage and claim the method gives
insight by, for instance, showing what intermediate elements or "hidden
units" enable a task to be done, but one suspects the avoidance of detailed
analysis is nonetheless a factor in the popularity of the method.

The Materials Are Very Different

The objection that the materials are totally different need not be taken
very seriously since, whatever the materials are, they still obey the rules of
physics and chemistry. It just means more surprises for the neuroscientist,
since he will come across unexpected properties and methods; birds do not
use propellers or jet engines, but they do obey the same laws of aero-
dynamics as machines that fly.
Perhaps one should be a little more cautious when it comes to computers and the brain, for the architecture as well as the materials are so different, but surely some at least of the problems encountered when performing a task on a serial computer will carry over to the performance of the same task by the brain. As long as that is so, we can learn from the analogy.

Minds and Machines Differ in Origin

Does it matter that brains are the product of millions of years of genetic evolution combined with a few months’ ontogenesis and several years of teaching and experience, whereas machines are designed by those brains and made by human hand? I can’t see why it should, and it could be claimed that the explicit knowledge required for design and construction gives deep insight into the nature of the mind. But once again there are cautions.

Evolution is an unprincipled and conservative designer. As a result, one rarely finds neat theoretical solutions embodied in brains, and in particular they often use parallel mechanisms for achieving a single goal. This fact tends to make life difficult for the experimentalist, because it frustrates attempts to do simple and effective controls. For instance, one might naively expect that if binocular vision enables people to judge distance, then blocking one eye would seriously impair this capacity, but motion parallax, knowledge of the normal sizes of objects, and other cues leave good distance judgment in one-eyed people. The use of alternative sensory cues for balance can cause confusion (see below), and the presence of many parallel methods has caused untold difficulty in the analysis of homing and other navigational feats in animals. It seems to be hard for us to avoid the automatic assumption that just one method is used to perform some difficult task, and this may partly result from the metaphor of the mechanical mind, for multiple alternative methods are rare in machines.

WHAT THE MIND DOES

Many of the things the mind and brain do are also done by machines, and when this is the case understanding the machine obviously helps. To understand how a man balances on one leg, one should understand the principles of servo-feedback, though to reinforce the point made above, one should also know that a man’s ability to stand on one leg with his eyes shut does not prove that vision is irrelevant to the task, and the ability to remain upright after the destruction of his vestibular organs does not prove they are unimportant either. But the brain also does things that no machine does, and here we get into trouble.

Things No Computer Does

To start with, consider the fact that we constantly model the inanimate world around us, and also monitor it for changes. We know thoroughly the route from home to office, and notice when a section of road is being repaired, or when the office door is newly painted. Much of this model-making is quite automatic and unconscious and we only know it has been done when something changes; we do not consciously monitor the spectral composition of sunlight or the pitch of the front door bell, but it’s pretty certain we would notice if they changed.

Robots are of course beginning to record and model their environments in order to find their way around in it, but compared with us they are extraordinarily backward, and so far they have little to teach us. But it would be premature to say that, for this reason, the metaphor is wrong or misleading, because the need for modeling has only recently arisen. It is likely that the principles will soon be better understood, and when this happens they may give us new insight into the brain’s methods of building useful models of the environment.

Modeling People

More serious problems arise when one considers the human brain’s propensity to model people. This starts at a very early age, and parents quickly realize that they are not the only ones trying to run the household; a baby quickly becomes pretty expert at parent-control. It seems to me that this really is a most un-machine-like process, and likely to remain so. One reason is that machines are designed to do something useful for their owners, whereas babies are not. No doubt one could incorporate a few tricks in a computer that would enable it to get the better of its user—in fact many of them appear to do this effortlessly, without deliberate design; but in such cases it is clear upon reflection that they are not really getting the better of us, they are simply failing to give the desired service. To improve the man–machine interface, an engineer might program a computer to learn about the user’s behavior, just as a baby learns about its parents’ behavior; but there the analogy ends, for the engineer’s purpose would always be to diminish the conflict of wills, while that is not the baby’s purpose at all. A machine is by definition intended to be useful, and therefore it must be designed so that its user’s will is unopposed as far as possible.

Here, then, we have reached a point where the machine is not a good metaphor for the mind, simply because no machine has been designed to do what the mind does. I am not saying there is a theoretical reason that they should not be so designed, only that engineers are not likely to
explore the means of doing something that it would be useless or counterproductive to do, and still less likely actually to do it. We, as neuroscientists, can explore the problems of one brain outguessing another, but we shall not at present get much help in doing so by regarding the mind as a machine, simply because that is not at present the sort of thing machines do.

Notice that this argument is only valid for the present; comparative and competitive interactions between computers can be modeled, and this may be very instructive for understanding human interactions. Hence the implications of the metaphor are not static, and it may mean something very different to future generations that have more understanding of the complexity of purely mechanistic interactions. All the same, for the present, the metaphor is, at best, unhelpful on such problems.

TAKING THE METAPHOR TOO SERIOUSLY

Perhaps this discussion has brought us to the reason for the unease and dislike aroused by the mind-machine metaphor. If we took it seriously, would we not regard other people’s minds as machines, and hence be interested in them primarily for their utility to us? Machines do not and should not oppose our wills, and nor should other people, according to the metaphor.

This would all be less worrying if the newspapers were not full of the deeds and misdeeds of individuals who behave as though the metaphor was their gospel, and such an attitude is not entirely unfamiliar even in academic circles. Of course the metaphor is not intended to be a moral exhortation saying: “Treat other people’s minds as machines that could be useful to you.” But when we use it aren’t we in effect saying: “It’s alright to treat minds as machines, because actually they are”?

Thus the metaphor may not be entirely innocuous, since it might influence how people think about minds and consequently how they treat other people. To follow up this thought we need to see how people actually use the idea of the mind.

Distinguishing Minds from Brains

So far, encouraged by the mind-machine metaphor, I have used the words “mind” and “brain” more or less interchangeably, but now we see that a distinction might be useful. The brain is what the metaphor properly applies to, while the mind is something different: it is the concept we use to describe the source of other people’s, and our own, behavior. Since the concept of a particular person’s mind is fashioned out of the observed behavior of that individual, it is a model one’s own brain makes of the

other individual’s brain. Thus minds are the brain’s models of itself and other brains, and the important thing is that the vast majority of people attribute behavior to such mind-models. As neuroscientists we believe it is the brain that controls behavior, but this is the belief and terminology of a small minority of experts; others attribute a person’s behavior to his mind and care nothing for the beautiful nerve cells that we dedicate our lives to. That may be a pity, but in justification of the majority’s attitude, pause to think how many of the facts you learn from this volume will alter the way you treat your colleagues, bring up your children, talk to the janitor, or vote at the next election. Your beliefs about other people’s minds, on the other hand, clearly do influence your way of life, and it is because the metaphor may modify your attitude to minds that it is potentially obnoxious.

One can be thoroughly mechanistic in believing that the brain controls behavior strictly within the laws of physics and chemistry without this distinction between mind and brain becoming a mockery. In most cases one has knowledge of only a minute fraction of the physical and chemical factors that are actually (we mechanistically believe) controlling the brain’s output, but this does not stop us from making quite good and reliable judgments about the actions our own and other minds will initiate. There is nothing unusual in a model having such predictive power in spite of its use of incomplete data; in fact, good models stringently select the data they represent, both in the case of one’s mental models of the physical environment with the people in it, and in the case of accurate scientific theories such as thermodynamics. The predictive power of a model depends on its correct identification of the dominant controlling factors and their influence, not upon its completeness. An incomplete model is often more generally useful than a more accurate one, as for example with Newtonian laws of motion.

Minds Take Over the Control of Behavior

Figure 1 is an attempt to persuade neuroscientists that minds are actually more important than brains, and to show how this comes about. In stage 1, the mechanistic brain is in sole command, but brain A observes other brains, B for example, also controlling their own behavior. Occasionally they interact, or their behaviors conflict, so brain A builds a model of the way brain B controls what B says and does, based on what A has seen and heard. This model of brain B inside the brain of A is B’s mind, shown in step 2; brain A uses his idea of B’s mind to predict what B will say and do, and this should be beneficial to A when living in the same environment as B. B of course has done the same and now has its internal model of A’s brain, A’s mind.
At this stage brain A can predict B's solo behavior, but this will not enable him to predict the outcome of interactions between A and B, for these cannot be modeled, since there is nothing inside A's brain suitable for B's mind to interact with. Brain A needs to make a model of the way its own brain generates its own behavior; this is A's mind, and the interactions between A's mind and B's mind allow brain A to model the interactions between A and B. By this time the two minds in brain A have become dominant in controlling much of the behavior of A, especially the verbal and other exchanges with other individuals. On the other hand, nonverbal and nonsocial behavior may still be generated predominantly by brain A, uninfluenced by the minds it has created within itself; these are, as we would say, mindless actions.

Of course all the time similar processes will have been going on in brain B, so A's models will not have it all their own way when it comes to making accurate predictions. Furthermore, if these mind-models are realistic, they should include minds within themselves; B's mind in brain A should, for example, have a little A's mind inside itself. For by this stage brain B will have its own version of A's mind, as shown in the right half of the figure. This infinite regress can presumably be cut short after a few cycles without seriously affecting the predictive power of the models, but it makes the brain into something more like a hall of mirrors than any currently understood machine. Surely, however, this hall of mirrors accurately portrays the fact that two people's minds interact with each other in an extraordinarily complex and intimate manner. At the very least, the interactions that this figure attempts to portray will have to be taken into account in any mechanistic accounts of higher human behavior, and I do not see how this can be done without giving the concept of the mind a very dominant role.

Why Minds are More Interesting

Now that we have a distinction between mind and brain, we can see why most people are much more interested in one than the other. The mathematician, linguist, historian, or literary critic is not concerned with the nuts and bolts of brain mechanisms, but with mind—the abstract model that appears to be responsible for the works that he studies and perhaps produces himself. It is the same with cars, domestic appliances, word processors, and so on: except for the expert, people simply do not care how they work as long as they do the job expected. Of course when minds fail to work as expected we may call in the brain specialist, but until then it is minds we deal with.

I think this distinction between the mechanistic brain and the mind-models it makes has interesting implications for our understanding of pain,
pleasure, ethics, and consciousness itself. As I have argued elsewhere, the
part of the brain's functioning of which we are consciously aware seems
to be confined to the inputs, outputs, and interactions of the minds of
Figure 1. Just as ordinary people are more interested in minds than brains,
so should be philosophers, for it is the source of peoples' actions rather than
their mechanism that is their primary concern. Minds are the conceptual
sources of this behavior, and although minds are the product of a mecha-
nicist brain, their actions and interactions are not going to be under-
stood in terms of physics and chemistry alone, even if they are ultimately
determined by physics and chemistry alone.

Notice that within this entirely mechanistic framework one can ask
interesting questions about minds. How is the stability of social interaction
affected by how good a model of B's brain A has in the form of B's mind?
Are inaccurate models unethical? Or might it be unethical to have too
good a model? How about the same questions applied to A's mind, the
model of his own brain? How far will taught precepts affect the form of
these models? Do we also have mind-models of deities? Surely these are
the sort of questions appropriate for theoretical analysis by philosophers
and others, and the mechanistic nature of the brain hardly affects them.
It is worth adding that such ideas might also form the basis for experiments.

Thus, to neuroscientists, the machine-like aspects of the brain are
immediately important, because they firmly direct our attention to the
physics and chemistry of the brain. But the metaphor should not be taken
to imply that there is no more to the brain than can be described in terms
of physics and chemistry, and the concept of mind is certain to be important
in any satisfactory account of the way the brain controls social behavior.
For this reason, philosophers and ordinary people are properly more con-
cerned with minds than brains.

CONCLUSIONS

To conclude, for most people in most circumstances the physical and
chemical causation of the brain's output is not as important as the observ-
able behavior that actually occurs; minds are what we attribute this
behavior to, and our common-sense understanding of them will almost
always give more useful predictions than knowledge of the physical and
chemical processes that underlie them. For neuroscientists trying to refine
such mechanistic accounts, the metaphor is appropriate and encouraging,
but the dislike it arouses in others seems fully justified by its implication
that other people's minds are mere machines and can be treated as such.
What is needed to prevent the metaphor of the mechanical mind san-
donishing unbridled egocentric behavior is to get rid of the prejudice that
machines are essentially simple and deterministic, and to gain an appreci-
ation of the complexity and difficulties in predicting behavior produced by
two or more minds interacting in the manner shown in Figure 1. There is no
reason to doubt that the brain is entirely mechanical, but it is a wonderful
 mechanism that can generate and use the concept of mind.

ACKNOWLEDGEMENTS

I would like to acknowledge considerable help toward clarifying these
ideas from Ian Glynn, Graeme Michison, Kathy Mullen, and Miranda
Weston-Smith.