Truth Journal

MINDS ARE SIMPLY WHAT BRAINS DO

Marvin Minsky
Massachusetts Institute of Technology

We all believe that we have minds - and that minds, whatever they may be, are not like other worldly things. What makes us think that thoughts are made of different stuff? Because, it seems, thoughts can't be things; they have no weights or sounds or shapes, and cannot be touched or heard or seen. In order to explain all this, most thinkers of the past believed that feelings, concepts, and ideas must exist in a separate mental world. But this raises too many questions. What links our concept about, say, a cat with an actual cat in the physical world? How does a cause in either world affect what takes place in the other world? In the physical world we make new things by rearranging other things; is that how new ideas come to be, or were they somewhere all along? Are minds peculiar entities, possessed alone by brains like ours - or could such qualities be shared, to different degrees, by everything? It seems to me that the dual-world scheme creates a maze of mysteries that leads to problems worse than before.

We've heard a good deal of discussion about the idea that the brain is the bridge between those worlds. At first this seems appealing but it soon leads to yet worse problems in philosophy. I maintain that all the trouble stems from making a single great mistake. Brains and minds are not different at all; they do not exist in separate worlds; they are simply different points of view--ways of describing the very same things. Once we see how this is so, that famous problem of mind and brain will scarcely seem a problem at all, because ...

Minds are simply what brains do.

I don't mean to say that brains or minds are simple; brains are immensely complex machines-and so are what they do. I merely mean to say that the nature of their relationship is simple. Whenever we speak about a mind, we're referring to the processes that move our brains from state to state. Naturally, we cannot expect to find any compact description to cover every detail of all the processes in a human brain, because that would involve the details of the architectures of perhaps a hundred different sorts of computers, interconnected by thousands of specialized bundles of connections. It is an immensely complex matter of engineering. Nevertheless, when the mind is regarded, in principle, in terms of what the brain may do, many questions that are usually considered to be philosophical can now be recognized as merely psychological-because the long-sought connections between mind and brain do not involve two separate worlds, but merely relate two points of view.

Memory and Change

What do brains do? Doing means changing. Whenever we learn or 'change our minds', our brains are engaged in changing their states. To comprehend the relationship between mind and brain, we must understand the relationship between what things do and what things are; what something does is simply an aspect of that thing considered over some span of time.

When we see a ball roll down a hill, we appreciate that the rolling is neither the ball itself, nor something apart in some other world - but merely an aspect of the ball's extension in space-time; it is a description of the ball, over time, seen from the viewpoint of physical laws. Why is it so much harder to appreciate that thinking is an aspect of the brain, that also could be described, in principle, in terms of the self-same physical laws? The answer is that minds do not seem physical to us because we know so little of the processes inside brains.

We can only describe how something changes by contrast with what remains the same. Consider how we use expressions like "I remember X." Memories must be involved with a record of changes in our brains, but such changes must be rather small because to undergo too large a change is to lose any sense of identity. This intrusion of a sense of self makes the subject of memory difficult; we like to think of ourselves as remaining unchanged - no matter how much we change what we think. For example, we tend to talk about remembering events (or learning facts, or acquiring skills) as though there were a clear separation between what we call the Self and what we regard as like data that are separate from but accessible to the self. However, it is hard to draw the boundary between a mind and what that mind may think about and this is another aspect of brains that makes them seem different to us from machines. We are used to thinking about machines in terms of how they affect other materials. But it makes little sense to think of brains as though they manufacture thoughts the way that factories makes cars because brains, like computers, are largely engaged in processes that change themselves . Whenever a brain makes a memory, this alters what that brain may later do.

Our experience with computers over the past few decades has helped us to clarify our understanding of such matters. The early applications of computers usually maintained a rather clear distinction between the program and the data on which it operates. But once we started to develop programs that changed themselves, we also began to understand that there is no fundamental difference between acquiring new data and acquiring new processes. Such distinctions turned out to be not absolute, but relative to other issues of perspective and complexity. When we say that minds are what brains do, we must also ask whether every other process has some corresponding sort of mind. One reply might be that this is merely a matter of degree: people have well-developed minds, while bricks or stones have almost none. Another reply might try to insist that only a person can have a mind -and, maybe, certain animals. But neither side would be wrong or right; the issue is not about a fact, but about when to use a certain word. Those who wish to use the term "mind" only for certain processes should specify which processes. The problem with this is that we don't yet have adequate ways to classify processes. Human brains are uniquely complex, and do things that no other things do - and we must try to learn how brains do those things.

This brings us back to what it means to talk about what something does. Is that different from the thing itself? Again it is a matter of how we describe it. What complicates that problem for common sense psychology is that we feel compelled to think in terms of Selves, and of what those Selves proceed to think about. To make this into a useful technical distinction, we need some basis for dividing the brain into parts that change quickly and parts that change slowly. The trouble is that we don't yet know enough about the brain to make such distinctions properly. In any case, if we agree that minds are simply what brains do, it makes no further sense to ask how minds do what they do.

Embodiments of Minds

One reason why the mind-brain problem has always seemed mysterious is that minds seem to us so separate from their physical embodiments. Why do we find it so easy to imagine the same mind being moved to a different body or brain - or even existing by itself? One reason could be that concerns about minds are mainly concerns about changes in states - and these do not often have much to do with the natures of those states themselves. From a functional or procedural viewpoint, we often care only about how each agent changes state in response to the actions upon it of other agents. This is why we so often can discuss the organization of a community without much concern for the physical constitution of its members. It is the same inside a computer; it is only signals representing changes that matter, whereas we have no reason to be concerned with properties that do not change. Consider that it is just those properties of physical objects that change the least - such as their colors, sizes, weights, or shapes - that, naturally, are the easiest to sense. Yet these, precisely because they don't change, are the ones that matter least of all, in computational processes. So naturally minds seem detached from the physical. In regard to mental processes, it matters not what the parts of brains are; it only matters what they do--and what they are connected to.

A related reason why the mind-brain problem seems hard is that we all believe in having a Self - some sort of compact, pointlike entity that somehow knows what's happening throughout a vast and complex mind. It seems to us that this entity persists through our lives in spite of change. This feeling manifests itself when we say "I think" rather than "thinking is happening", or when we agree that "I think therefore I am," instead of "I think, therefore I change". Even when we recognize that memories must change our minds, we feel that something else stays fixed - the thing that has those memories. In chapter 4 of *The Society of Mind*[l] I argue that this sense of having a Self is an elaborately constructed illusion - albeit one of great and practical value. Our brains are endowed with machinery destined to develop persistent self-images and to maintain their coherence in the face of continuous change. But those changes are substantial, too; your adult mind is not very like the one mind you had in infancy. To be sure, you may have changed much since childhood - but if one succeeds, in later life, to manage to avoid much growth, that poses no great mystery.

We tend to think about reasoning as though it were something quite apart from the knowledge and memories that it exploits. If we're told that Tweety is a bird, and that any bird should be able to fly, then it seems to us quite evident that Tweety should be able to fly. This ability to draw conclusions seems (to adults) so separate from the things we learn that it seems inherent in having a mind. Yet over the past half century, research in child psychology has taught us to distrust such beliefs. Very young children do not find adult logic to be so self evident. On the contrary, the experiments of Jean Piaget and others have shown that our reasoning abilities evolve through various stages. Perhaps it is because we forget how hard these were to learn that they now appear so obvious. Why do we have such an amnesia about learning to reason and to remember? Perhaps because those very processes are involved in how we remember in later life. Then, naturally, it would be hard to remember what it was like to be without reason - or what it was like to learn such things. Whether we learn them or are born with them, our reasoning processes somehow become embodied in the structures of our brains. We all know how our logic can fail when the brain is deranged by exhaustion, intoxication or injury; in any case, the more complex situations get, the more we're prone to making mistakes. If logic were somehow inherent in Mind, it would be hard to explain how things ever go wrong but this is exactly what one would expect from what happens inside any real machine.

Freedom of Will

We all believe in possessing a self from which we choose what we shall do. But this conflicts with the scientific view that all events in the universe depend on either random chance or on deterministic laws. What makes us yearn for a third alternative? There are powerful social advantages in evolving such beliefs. They support our sense of personal responsibility, and thus help us justify moral codes that maintain order among the tribe. Unless we believed in choice-making entities, nothing would bear any credit or blame. Believing in the freedom of will also brings psychological advantages; it helps us to be satisfied with our limited abilities to make predictions about ourselves - without having to take into account all the unknown details of our complex machinery. Indeed, I maintain that our decisions seem "free" at just the times at which what we do depends upon unconscious lower level processes of which our higher levels are unaware - that is, when we do not sense, inside ourselves, any details of the processes that moved us in one direction or the other. We say that this is freedom of will, yet, really, when we make such a choice, it would be better to call it an act of won't. This is because, as I'll argue below, it amounts to terminating thought and letting stand whatever choice the rest of the mind already has made.

To see an example of how this works, imagine choosing between two homes, one of which offers a mountain-view, while the other is closer to where you work. There is no particularly natural way to compare such unrelated things. One of the mental processes that are likely to become engaged might be constructing a sort of hallucination of living in that house, and then reacting to that imaginary episode. Another process might imagine a long drive to work, and then reacting to that. Yet one more process might then attempt to compare those two reactions by exploiting some memory traces of those simulations. How, then, might you finally decide? In one type of scenario, the comparison of the two descriptions may seem sufficiently logical or rational that the decision seems to be no mystery. In such a case we might have the sense of having found a "compelling reason"--and feel no need to regard that choice as being peculiarly free.

In another type of scenario, no such compelling reason appears. Then the process can go on to engage more and more mechanisms at increasingly lower levels, until it engages processes involving billions of brain cells. Naturally, your higher level agencies - such as those involved with verbal expressions--will know virtually nothing about such activities, except that they are consuming time. If no compelling basis emerges upon which to base a definite choice, the process might threaten to go on forever. However, that doesn't happen in a balanced mind because there will always be other, competing demands from other agencies. Eventually some other agency will intervene - perhaps one of a supervisory character[2] whose job it is to be concerned, not with the details of what is being decided, but with some other economic aspect of the other systems' activities. When this is what terminates the decision process, and the rest is left to adopt whichever alternative presently emerges from their interrupted activities, our higher level agencies will have no reasonable explanation of how the decision was made. In such a case, if we are compelled to explain what was done, then, by default, we usually say something like "I decided to.'[3] This, I submit, is the type of situation in which we speak of freedom of choice. But such expressions refer less to the processes which actually make our decisions than to the systems which intervene to halt those processes. Freedom of will is less to do with how we think than with how we stop thinking.

Uncertainty and Stability

What connects the mind to the world? This problem has always caused conflicts between physics, psychology, and religion. In the world of Newton's mechanical laws, every event was

entirely caused by what had happened earlier. There was simply no room for anything else. Yet common sense psychology said that events in the world were affected by minds: people could decide what occurred by using their freedom of will. Most religions concurred in this, although some preferred to believe in schemes involving divine predestination. Most theories in psychology were designed to support deterministic schemes, but those theories were usually too weak to explain enough of what happens in brains. In any case, neither physical nor psychological determinism left a place for the freedom of will.

The situation appeared to change when, early in this century, some physicists began to speculate that the uncertainty principle of quantum mechanics left room for the freedom of will. What attracted those physicists to such views? As I see it, they still believed in freedom of will as well as in quantum uncertainty--and these subjects had one thing in common: they both confounded those scientists' conceptions of causality. But I see no merit in that idea because probabilistic uncertainty offers no genuine freedom, but merely adds a capricious master to one that is based on lawful rules.

Nonetheless, quantum uncertainty does indeed play a critical role in the function of brain. However, this role is neither concerned with trans-world connections nor with freedom of will. Instead, and paradoxically, it is just those quantized atomic states that enable us to have certainty! This may surprise those who have heard that Newton's laws were replaced by ones in which such fundamental quantities as location, speed, and even time, are separately indeterminate. But although those statements are basically right, their implications are not what they seem - but almost exactly the opposite. For it was the planetary orbits of classical mechanics that were truly undependable - whereas the atomic orbits of quantum mechanics are much more predictably reliable. To explain this, let us compare a system of planets orbiting a star, in accord with the laws of classical mechanics, with a system of electrons orbiting an atomic nucleus, in accord with quantum mechanical laws. Each consists of a central mass with a number of orbiting satellites. However, there are fundamental differences. In a solar system, each planet could be initially placed at any point, and with any speed; then those orbits would proceed to change. Each planet would continually interact with all the others by exchanging momentum. Eventually, a large planet like Jupiter might even transfer enough energy to hurl the Earth into outer space. The situation is even less stable when two such systems interact; then all the orbits will so be disturbed that even the largest of planets may leave. It is a great irony that so much chaos was inherent in the old, deterministic laws. No stable structures could have evolved from a universe in which everything was constantly perturbed by everything else. If the particles of our universe were constrained only by Newton's laws, there could exist no well defined molecules, but only drifting, featureless clouds. Our parents would pass on no precious genes; our bodies would have no separate cells; there would not be any animals at all, with nerves, synapses, and memories.

In contrast, chemical atoms are actually extremely stable because their electrons are constrained by quantum laws to occupy only certain separate levels of energy and momentum. Consequently, except when the temperature is very high, an atomic system can retain the same state for decillions of years, with no change whatever. Furthermore, combinations of atoms can combine to form configurations, called molecules, that are also confined to have definite states. Although those systems can change suddenly and unpredictably, those events may not happen for billions of years during which there is absolutely no change at all. Our stability comes from those quantum fields, by which everything is locked into place, except during moments of clean, sudden change. It is only because of quantum laws that what we

call things exist at all, or that we have genes to specify brains in which memories can be maintained - so that we can have our illusions of will.[4]

QUESTIONS

Question: Can you discuss the possible relevance of artificial intelligence in dealing with this conference?

Artificial intelligence and its predecessor, cybernetics, have given us a new view of the world in general and of machines in particular. In previous times, if someone said that a human brain is just a machine, what would that have meant to the average person? It would have seemed to imply that a person must be something like a locomotive or a typewriter. This is because, in earlier days, the word machine was applied only to things that were simple and completely comprehensible. Until the past half century - starting with the work of Kurt Goedel and Alan Turing in the 1930s and of Warren McCulloch and Walter Pitts a decade later - we had never conceived of the possible ranges of computational processes. The situation is different today, not only because of those new theories, but also because we now can actually build and use machines that have thousands of millions of parts. This experience has changed our view. It is only partly that artificial intelligence has produced machines that do things that resemble thinking. It is also that we can see that our old ideas about the limitations of machines were not well founded. We have learned much more about how little we know about such matters.

I recently started to use a personal computer whose memory disk had arrived equipped with millions of words of programs and instructive text. It is not difficult to understand how the basic hardware of this computer works. But it would surely take months, and possibly years, to understand in all detail the huge mass of descriptions recorded in that memory. Every day, while I am typing instructions to this machine, screens full of unfamiliar text appear. The other day, I typed the command "Lisp Explorer", and on the screen appeared an index to some three hundred pages of lectures about how to use, with this machine, a particular version of LISP, the computer language most used for research in artificial intelligence. The lectures were composed by a former student of mine, Patrick Winston, and I had no idea that they were in there. Suddenly there emerged, from what one might have expected to be nothing more than a reasonably simple machine, an entire heritage of records not only of a quarter century of technical work on the part of many friends and students, but also the unmistakable traces of their personalities.

In the old days, to say that a person is like a machine was like suggesting that a person is like a paper clip. Naturally it was insulting to be called any such simple thing. Today, the concept of machine no longer implies triviality. The genetic machines inside our cells contain billions of units of DNA that embody the accumulated experience of a billion years of evolutionary search. Those are systems we can respect; they are more complex than anything that anyone has ever understood. We need not lose our self-respect when someone describes us as machines; we should consider it wonderful that what we are and what we do depends upon a billion parts. As for more traditional views, I find it demeaning to be told that all the things that I can do depend on some structureless spirit or soul. It seems wrong to attribute very much to anything without enough parts. I feel the same discomfort when being told that virtues depend on the grace of some god, instead of on structures that grew from the honest work of searching, learning, and remembering. I think those tables should be turned; one ought to feel insulted when accused of being not a machine. Rather than depending upon

some single, sourceless source, I much prefer the adventurous view of being made of a trillion parts--not working for some single cause, but interminably engaged in resolving old conflicts while engaging new ones. I see such conflicts in Professor Eccles' view: in his mind are one set of ideas about the mind, and a different set of ideas that have led him to discover wonderful things about how synapses work. But he himself is still in conflict. He cannot believe that billions of cells and trillions of synapses could do enough. He wants to have yet one more part, the mind. What goodness is that extra part for? Why be so greedy that a trillion parts will not suffice? Why must there be a trillion and one?

Eccles: I am being completely misrepresented. There is no evidence to quote from me in favor of one of those things.

Minsky: I am glad to hear that.

Eccles: I do not look at the mind as an additional point or anything like that. The mind is an entity more complex really than the brain.

Minsky: I did not realize that you thought the mind has many parts.

Eccles: Have you ever seen a diagram where I show all those aspects of the mind ... imagining, . All those feelings are aspects of the mind.

Minsky: But why are not they aspects of the brain?

Eccles: Well, they are related to the brain but I think that you will find...

Minsky: O.K., I did not realize you had that complex a theory. I stand corrected.

Ayer: Can we abandon discussion on this issue.

Minsky: I think that Searle's argument is wrong (I don't know how many listeners are familiar with it) in maintaining that there is a difference between genuine understanding and simulated understanding. I do not think that there is any such thing as genuine understanding. A Martian, or some alien machine that had a trillion trillions of parts might consider us to be simple machines without any genuine understanding--because of being a trillion times simpler than them. They might just regard us as very cute toys.

[**comment:** I think you are over impressed by size.]

All I can say is I think that one of the serious problems of our time - and of philosophy in particular--is not realizing that size can be terribly important. There are important differences between a machine with a trillion synapses and one that has only seven. Machines with very few parts cannot think. In this respect, large size is at least a necessary condition.

Lewis: You say that minds are what brains do ...

Minsky: When I say that minds are what brains do, that doesn't mean that they know what they do. You don't know why you are thirsty, for example. It could be because some internal water regulating mechanism has crossed a certain threshold of activation. It could be what those who study the behavior of animals call "displacement"-in which conflicting activities in

other brain centers have aroused the brain center for thirst. There are many possible causes for wanting a glass of water. We like to think we are self aware in the sense of knowing what happens inside our minds. But we don't really understand our minds any more than we understand our brains. We do not have reliable insights into our own psychologies. We do not know how we see what we see. When you ask people about their beliefs and then discuss their replies, you often find them changing their minds about what they said. We talk about knowing and believing, but it seems to me that those ideas are only first approximations. The different parts of a person's mind may maintain different views about the world. It seems to me that philosophers are naive about psychology in assuming that certain things are known or believed by the entire person - rather than by various parts of the person.

Ayer: Some philosophers.

Minsky: Some, I think, most philosophers. Perhaps with the exception of Daniel Dennett and a few others. Most of what I read about philosophy seems based on naive ideas about psychology. Philosophy often has to change when, after it raises questions, science starts to answer them.

Notes

1. Marvin Minsky, The Society of Mind, Simon and Schuster, 1987; Heinemann & Co., 1987.
2. The idea of supervisory agencies is discussed in section [6.4] of [1].
3. In 22.7 of [1] I postulate that our brains are genetically predisposed to compel us to try to assign some cause or purpose to every change - including ones that occur inside our brains. This is because the mechanisms (called trans-frames) that are used for representing change are built automatically to assign a cause by default if no explicit one is provided.
4. This text is not the same as my informal talk at the conference. I revised it to be more consistent with the terminology in [1].