Ecclesiastes:
A Report from the Battlefields of the Mind-Body Problem

Stanley Letovsky

On 1-3 March I had the privilege of being present at a most rare and interesting event. A conference was held at Yale University entitled "Artificial Intelligence and the Human Mind: An International, Interdisciplinary Symposium." I am a graduate student in AI at Yale and was surprised, along with my colleagues, to learn of this conference just two days before it began. The program included some very famous and impressive names, such as Sir John Eccles, the Nobel laureate neurobiologist; the physicists Henry Margenau and Eugene Wigner; Marvin Minsky and Michael Arbib, both eminent AI researchers; and Hans Moravec and Doug Lenat, two younger stars of AI. The conference was being organized by two agencies outside the usual AI circles: Truth (a journal of modern thought) and The International Institute for Mankind.1

It began on a Saturday morning, and I was hard pressed to even find out the location of the conference on campus, so minimal was our information about it. On the way in I ran into Minsky, which is rather like a 10-year-old boy running into his favorite World Wrestling Federation star on the street. We talked briefly as we went to the lecture hall.

The conference began presently, with a gentleman standing up to introduce someone who was going to stand up to introduce the first speaker. In his southern accent, this fellow made some remarks about one of the sponsoring organizations, Truth magazine, whose members are honest, open-minded seekers of truth, and that part of this truth is the historical person of Jesus Christ, who should not be confused with the representations of him put forward over the years by various organized religions. I sat there shocked, along with my buddies from the AI lab; these comments are not the sort of thing one is used to hearing at a colloquium. It appeared that this conference had been put together by a group of southern evangelical Christians.

Our discomfort was short lived; he sat down, and Sir John Eccles got up to introduce the first main speaker. This speaker was to be the physicist Henry Margenau, in whose honor the conference was being given. Now in or near his eighties, his career had been devoted to quantum mechanics. Eccles referred to Margenau's important contributions to the philosophy of science and mentioned several books. I had never heard of him, personally. Margenau then got up and began talking, slowly and ponderously, summarizing in a seemingly random fashion the development of various physical theories during his lifetime. He related interesting stories about the opinions of the quantum-generation physicists, such as Bohr and Einstein, about the mind and the universe, but mostly he was presenting familiar theories I wandered into the hall where I ran into an equally bored Marvin Minsky and chatted with him about my semiwhimsical idea for a computational theology. This idea is that if, as AI would have it, mind is a kind of computational process, then perhaps the computational processes of nature, such as evolution, support the minds of divine beings. I said you needed a kind of IQ test that you could apply to any process at all to decide if it was intelligent, and we discussed what such a test would have to be like. Minsky thought it would have to be sensitive to the amount of computation a pro-
cess performed in order to solve a problem of a given size I disagreed because of my bias that evolution should pass the test and argued it didn't matter how much resources a process used to solve a problem, only whether it could.

Then we went back to the lecture, just in time to hear some follow-up discussion after Margenau's talk, including Sir John singing the praises of the wonderful theory that Margenau had just outlined, which he was sure would lay the basis for a revolutionary new understanding of mind and brain I wondered what Eccles could be talking about, having heard only old theories from Margenau before I left the room.

I went to lunch with some other AI graduate students, and there I learned that Margenau, near the end of his talk, advanced the theory that the mind acts on the brain by selecting among the possibilities permitted by quantum mechanics in such a way that the probabilistic predictions of quantum mechanics were still respected. For me, this theory was at once obvious and obviously wrong.

It was obvious because of events that took place in science at the beginning of this century. During the several hundred years interval when the mechanical theories of Isaac Newton dominated physics, philosophers were deeply disturbed by one of the implications of these theories. This implication was that the universe was essentially a deterministic machine acting out a history which was precisely ordained by its initial conditions; so, if you knew the state of the universe in any single instant, you could in principle calculate all past and future states. Philosophers found this conclusion offensive because it seemed to leave no room for our intuitive sense of our own free will.

When Newton's theories were superceded early in this century by the theories of relativity and quantum mechanics, this picture changed. Quantum theory asserted that at the level of fundamental particles, physics was not and in principle could not be deterministic; particles obeyed probabilistic laws only; they moved along probabilistic paths that defied human intuitions, and their detailed movements could never be known with certainty. Philosophers leaped on this fact as the solution to the problem of free will; because the universe was unpredictable, free will was again possible.

By the time I became a thinking scientist, it was apparent to most people who had thought about the problem at all that quantum indeterminacy could have a bearing on human behavior only if the probabilistic behavior of the submicroscopic particles was somehow coupled to the pattern of firing of neurons in the brain. Such a coupling was not entirely implausible, but it seemed of little use philosophically because it implied not so much free will as a deterministic machine with a quantum random number generator influencing its behavior. Moreover, schooled as I was amid the exciting developments of biology and computer science in the 1970s and 1980s, there seemed little reason to doubt that a neuronal analysis of the brain and a computational analysis of the mind would together suffice to enable us to understand both brain and mind, and there was, therefore, no need to worry about quantum effects.

Nonetheless, here was Margenau, an eminent physicist, professing exactly this silly theory that no thinking person could possibly take seriously. In his version, free will was salvaged because the mind was a separate nonphysical entity that controlled the body by influencing the firing of the neurons in a way which was allowed by quantum mechanics. Because the mind only selected among the possibilities permitted by quantum mechanics, it needed no energy to function, and because it did not alter the statistical behavior of particles, it did not violate the laws of quantum mechanics, or equivalently, its effects could never be physically demonstrated. (This interpretation is mine; it might not be his.) As a theory, it is elegant, but it struck (and strikes) me fundamentally as a kind of mysticism, an attempt to slide a rather standard solist view of the mind into a sort of loophole in physics. And Sir John Eccles, a scientist of tremendous stature, was endorsing this view! Incredible.

After lunch, the conference resumed. Memory and time preclude me from describing each talk in as much detail as it warrants. Minsky gave a talk representing the AI position on the mind; "minds are what brains do," was his motto. He spoke without notes, seeming to make it up as he went along (in contrast to an earlier presentation by an English philosopher who simply read his paper aloud, as those fellows often do). Minsky asserted that it was impossible to...
have thought without some kind of short-term memory device to keep track of what you are thinking about, and that this device would have to be some kind of physical system. He talked about the complexity of both brain and mind and how little of this complexity is accessible to conscious awareness. He chided those who insisted that self-awareness was a singular human quality which machines could never duplicate, the extent of our introspective powers was so small compared to the complexity of our thought processes that our self-awareness could be considered only marginal at best.

Minsky also said that not long ago, to say people were like machines would have been insulting because machines in those days were items such as typewriters and cars, which at most had a few hundred parts. Today, though, we have machines with millions of parts, and the claim becomes a little more tenable, but we should consider that we ourselves are machines with not billions but trillions of parts, counting only brain cells. He invited us to consider the dignity of being a machine with a trillion complex interacting parts. Was this not enough? Why was John Eccles so greedy that he couldn’t be satisfied with a trillion parts but must have a trillion and one, demanding this one extra part, a soul? Eccles became quite irate at this statement and blustered that he had said no such thing and that in his theory the mind was a complex object with many parts. Minsky apologized, and the talk continued.

The program was organized so that each talk was followed by a response from one of the people in the program, which in turn was followed by questions and speeches from the audience. Several of the talks, and many of the questions, made clear that there was a school of thought in the room which was skeptical of the prospects for a purely physical explanation of the mind.

At the end of this session, Hans Moravec, a robotics researcher, got up to deliver a response to a paper by a philosopher professing some form of soulist mysticism. Moravec, like Minsky, was driven by these presentations to present the AI view in as shocking and outrageous a manner as possible. In a playful yet serious tone, he put forward the claim that not only was it possible in principle for a machine to be conscious but that he had already constructed several conscious machines in his laboratory. He described one robot that among many other capabilities has a program which allows the robot to monitor its environment for precipices and direct itself to steer away from them. When it came near a precipice, this robot would turn around and scuttle away. According to Moravec, the robot’s behavior resembled and could properly be labeled, fear.

He went on to describe how he evolved robots in his laboratory by studying the behavior of different designs in different environments and incrementally making changes, preserving favorable parts of each design. I saw Eccles listening with appreciation, perhaps recognizing in Moravec’s enthusiasm and obvious love of his work the image of himself as a younger man, despite the ideological gulf.

Later during a break, I was chatting with Moravec when Eccles came over and started talking to Moravec about the need to save these early attempts at robotics in some kind of museum. They talked amiably about this idea, and Eccles reminisced about his experiences with the earliest computers and how invaluable they were in studying the nervous system. They seemed instant friends; across disciplines, generations, and ideological positions, they were both scientists who loved doing science.

That evening I mulled over what had gone on during the day. My strongest impression was of an ideological battleground, a debate between dualists, who believe the mind is somehow distinct from the brain, (Eccles, Margenau, and several others), and what for want of a better term I call monists, who believe that the mind can be explained entirely in terms of the activity of the brain. This latter camp included Minsky, Moravec, and Arbib as well as myself and my fellow students. The contrasts between the two groups were striking. The dualists were mostly British physicists and philosophers in their seventies and eighties; the monists were younger, mostly American in manners if not always in origin, and
He (Minsky) invited us to consider the dignity of being a machine with a trillion complex interacting parts.

Was this not enough?

technologically oriented. Somehow, the sponsors had contrived to assemble the biggest scientific guns they could find to support a dualist position and paired them off against the leading exponents of the opposing position. Apparently, they had done this sort of thing before, having sponsored a conference of eminent theists and atheists and another of evolutionists and creationists.

I was struck by the historic proportions of the debate and the personalities involved. It reminded me of those debates which occurred in England after Darwin’s theory first came out, where eminent scientists ridiculed the theory from the bedrock of commonsense prejudices. As far as I was concerned, the monists were far and away more convincing in their arguments and presentations, but I was no objective observer. The dualist papers seemed like pseudoscientific mysticism. In addition to Margenau’s theory of the quantum mind-brain interface, there was also a Kaluza-Klein theory of the mind. Kaluza-Klein theories are fashionable in physics these days; they explain the various forces of physics by postulating extra dimensions beyond the normal four of space-time in which the forces are represented geometrically in the curvature of the space, much like Einstein described gravity in four dimensions. Some of these theories propose as many as 27 dimensions, although most of them are claimed to curl up to almost negligible thickness. In any case, John Smythies put forward the hypothesis that the mind inhabited some of these extra dimensions or at least some analogous but as yet unsuspected dimensions. Again, a pretty fantasy was described, but not the slightest argument offered in support of it.

Later that evening I went with a friend to the hotel where the conferees were staying and found a number of them gathered at tables in a scheduled after dinner open conversation. I sat at a table with Minsky and Moravec, and we told stories and traded favorite scientific and science fiction ideas till late in the evening. It was heaven.

The next morning, there was a presentation by Sir John Eccles. He presented recent data on the changes in blood flow to different regions of the brain as human subjects performed different tasks. These blood-flow changes clearly showed that different regions of the brain worked harder in the performance of different tasks. There was also a temporal pattern to the activation of the regions, and it happens that in each task the same area was activated first, an area known to be responsible for voluntary motor actions. Because activity appeared first in this area, with no apparent prior cause, Eccles suggested that this was the site where the mind acted on the brain to initiate physical motions, using the quantum mechanical coupling espoused by Margenau.

Eccles’ keenness of mind was striking, especially considering his age, as was his apparent determination to formulate Margenau’s thesis in a coherent and scientifically testable manner. This approach was quite in contrast to the other dualists, who seemed intent on designing theories that could not be falsified. He presented figures which he claimed established the plausibility of Margenau’s theory by...
showing that the energy required to cause a vesicle of neurotransmitter which was already poised to fuse with a synaptic membrane to actually do so was small enough to conceivably be borrowed from the vacuum and returned within the time required by Heisenberg's uncertainty principle.

The reply to Eccles' paper was delivered by Michael Arbib, an extremely articulate monist with traces of an Australian accent and a veneer of British dignity barely concealing an impish sense of humor. He thanked the sponsors for creating this event and remarked chivalrously that the research required by Sir John in the further development of his dualist position was exactly the research which he wanted to see done in the furtherance of his own monist position; therefore, ample room for cooperation in research existed, even if the researchers must agree to disagree philosophically. He got in a cute feigned Freudian slip by referring at one point to the "interaction of monists and dualists" when you were expecting to hear 'mind and body.' He also pointed out that Eccles' blood-flow data were no more supportive of the dualist position than the monist position because of course monists would expect parts of the brain that were working harder to use more blood.

Also on Sunday was a presentation by the British mathematician John Lucas. He claimed that AI could never succeed and that a machine was, in principle, incapable of doing all a mind could do. His argument went like this. Any computing machine is essentially equivalent to a system of formal logic. The famous Gödel incompleteness theorem shows that for any formal system powerful enough to be interesting, there are truths which cannot be proved in that system. Because a person can see and recognize these truths, the person can transcend the limitations of the formal system. Because this statement is true of any formal system, a person can always transcend a formal system; therefore, a formal system can never be a model of a person.

Minsky gave the rebuttal to this claim; he said that formal systems have nothing to do with AI or the mind because formal systems require perfect consistency; whereas, AI requires machines that make mistakes, that guess, that learn, and evolve. I was less sure of this refutation; although I agreed with Minsky, I was worried that because the algorithms for doing all this guessing and learning and mistake making would run on a computer, there was still a level of description at which the AI model looked like a consistent formal system. This statement is equivalent to saying that your theory of the mind is a consistent theory. I was worried that Lucas could revive his argument at this level, and I wanted a convincing refutation. I worked on the problem all through lunch with two other students and eventually came up with a plausible refutation. Lucas had talked vaguely of formal systems, but when you put enough detail into the formal system to allow it to denote the kind of machine Minsky was talking about, the Gödel sentence (the assertion that cannot be proved within the system) for this formal system had nothing to do with the beliefs of the mind modeled by the system. The system could prove that the mind would believe the Gödel sentence, even if the system couldn't prove the Gödel sentence itself. I tried to explain this point to Lucas but couldn't fully get the point across before we were interrupted by the start of the afternoon session.

That afternoon Moravec gave his position paper. He described a technological thought experiment in which a person's mind is transferred from the brain to a computing device without the person ever losing consciousness. He said there were many ways you could imagine performing this transfer, but the simplest and most grisly was as follows: You pull a small piece of brain tissue slightly away from the brain without disconnecting the neurons in any way. Then you insert wires into the connections with a tangle arrangement so that you can electrically switch the fragment of tissue in and out of the rest of the brain circuitry. Then you analyze this segment's connectivity noninvasively in some way, and you construct an artificial component with identical input-output behavior. Next, you hook this component into the circuit and let the person toggle between the original and artificial components, asking if they notice any differences. (When in doubt, you might want to ask if the person's lawyer can detect any differences.) Once a satisfactory component has been installed, the piece of tissue can be dispensed with. The entire brain is then whittled away in this manner.

Moravec observed that if you just summarize the effects of the procedure, it sounds very dualistic; you have transferred a mind from one body to another—in effect, a mind transplant. However, it is all achieved within a monist framework.

He went on to discuss some other implications of the thought experiment. Because the mind was now in software, you could back it up on disk. If you knew you were going on a dangerous mission, you could save a copy of yourself. If you get killed, your friends could resurrect the copy; it would be you up to the point where you stored the mind, with a short gap in memory corresponding to the time spent on disk—not too high a price: a small, finite interval of death in exchange for immortality.

Minds embodied in machines would be able to perform other tricks as well, such as fissioning into identical copies that then diverge in history and identity. At what point are they distinct?
individuals? How far do they have to diverge before it would be considered murder if you killed one of them? The question was analogous, Moravec said, to the question of when a fetus becomes a person.

He closed with the following questions. Does it matter what kind of hardware the mind is running on as long as experience remains continuous? If not, does it matter whether experience is continuous as long as the software necessary for experience exists (for example, stored on disk), and you can have experience back at any time? If not, does it matter if the software exists as long as it might exist? All software already "exists" in some sense within the abstract world of mathematical objects and computer programs.

Lucas gave the reply, which was mainly appreciative, taking Moravec's visions in the humorous spirit that they were given. There seemed to be a range of reactions in the audience, from delighted to scandalized. At one point, someone asked Lucas if he would subject himself to Moravec's mind transplant in exchange for immortality. Lucas declined. Minsky would later jokingly refer to this attitude as "pro death."

I had invited Minsky, Moravec, and Arbib to dine with myself and some colleagues from the Yale AI lab; these were, after all, visiting dignitaries. And dine we did Sunday night, a party of 10 or so, at an expensive Italian restaurant across from their hotel in New Haven. The food was passable, the company a rare delight.

Next morning, the proceedings opened with remarks by a man who faintly resembled Jerry Falwell and who sermonized about how computers will never replace the love that we humans all need, God's love; he then began reading passages from the Bible. He kept this reading up for close to half an hour despite the fact that his chosen texts had no apparent relation to the topic of the conference. I occupied myself by passing notes back and forth with the student next to me, wherein we commented wittily on the situation, something I used to do a lot in high school. Somehow, the situation felt the same as in those days: you find yourself trapped in a chair by the constraints of civilized behavior, while someone stands at the front of the room and says things that mean nothing to you but boredom.

The first speaker was Arbib, who made some cutting remarks about how the sermonizer would not have said many of the things he said had he been present during the two preceding days. In reaction to his quoting of the Bible, Arbib said he would quote from another very good book called Sirens of Titan, by Kurt Vonnegut. The passage he quoted concerned the beings on the planet Tralfamadore, who happened to be sentient robots. According to their legends, they had been created by a race of biologically evolved organisms. These biological organisms were obsessed with finding out the meaning of existence, and they built intelligent machines to assist them in this task. The machines were very good at this task and soon found out that there was no meaning at all to the creatures' existence. The creatures became very upset at this news and began slaughtering themselves, and each other, and they got the machines to help. The machines turned out to be much better at this task too, and soon all the creatures were dead.

Having expressed himself on the subject of the Southerner's sermon, Arbib proceeded to the main part of his talk, which I am sorry to say I don't remember at all. The pointer is gone. I remember that it was eloquent and witty and monistic; that it contained the phrase "Ecclesiasts and Arbirians"; and that after responding to the last question, Arbib left the podium saying, "I have one final thing to tell you; I am not actually a human being; I am in fact a robot built by Marvin Minsky in 1968."

Later that morning, Doug Lenat gave his talk. He outlined what he sees as the major obstacles currently blocking the creation of an intelligent computer—the knowledge-acquisition bottleneck and the need to have a lot of knowledge already in the system before learning can occur with any speed. He then gave a summary of his own work in AI over the last 10 years and how it was attempting to address these problems. His work, which is among the most imaginative in AI, concerns discovery, creativity, and learning in machines. He is currently involved in a project to codify over a decade a tremendous body of commonsense knowledge in machineusable form. As usual, his presentation was polished, his slides cute. The audience was mesmerized. Afterward, the dualists were falling all over themselves to say that if this was AI they were all for it, and that Lenat's kind of AI was pure technology and constituted no threat to the uniqueness and sanctity of the human spirit. Eccles got up and virtually begged Lenat to put his talents to work in the study of the brain. Various AI people in the audience tried to get Lenat to take a stand on the relevance of his work to the mind-body problem, or the question of whether a computer intelligence could in principle do all that human intelligence does. He dodged the question, claiming that he was not concerned with modeling the mind; he was simply building intelligent technologies, and the question was of no interest to him. Pressed harder, he admitted that success in his efforts would be extremely useful to people working in cognitive modeling and vice versa.

After Lenat's talk, the most ancient looking and rickety dualist of them all got up to speak, and half the audience, including myself, bolted in less than fifteen seconds. I went for lunch, determined to miss the rest of the conference, which was scheduled to go on throughout the afternoon. There was no one in the afternoon session I
particularly wanted to hear. I went home and started fantasizing about presenting some of my own views to the audience at the conference. Of course, I had done just this in bits and pieces all along, but I was feeling an urge to make a summarizing statement on the whole thing. This urge stemmed partly from knowledge that the sponsors were taping the proceedings and having them transcribed; I wanted to see my words in a book alongside Eccles’ and Minsky’s, even if my name would be rendered only as “participant.” I wrote down some notes on paper, and then on impulse I returned to the conference, sermon in hand.

The afternoon session began with a paper by Brian Josephson, inventor of the Josephson junction, the semiconductor technology on which IBM based an abortive effort to develop superfast superconducting computers. He was a strange, small, shy British dualist with his own unique mysticism, his presentation was on meditation as a technique for studying the mind. ’Nuff said. With the AI stars almost all gone, and the conference in its last hours, the audience had dwindled to a small fraction of its peak size.

Professor David Martin gave a most interesting talk on the myths involved in shaping individual and collective responses to images such as the rational machine or the free-willed human being. He also talked about how science offers visions of technological utopias and dystopias that are strikingly like religious visions of heaven and hell. The analysis was literary, somewhat in the style of Marshall McLuhan. At the end, Sir Alfred Ayer, a renegade, elder, young-thinking, monist-leaning British philosopher, got up to give a response. Sir Alfred had misunderstood Martin to mean that AI would upset people and, therefore, shouldn’t be done; so he rushed into the breach with an outraged defense of scientific freedom.

Martin, genuinely confused, insisted that he had said nothing of the kind and apologized for miscommunicating his message.

In a lull in the question period afterward, I gave my sermon, reading from my notes like some pompous British philosopher. I reproduce them here:

I would like to reiterate the point I made earlier today: that the dualists’ fear that a scientific explanation of the mind will desacralize our experience is unwarranted. The impending scientific theory of the mind is not the first time in history that a religious or magical theory has had to give way before a scientific one. There was the triumph of Copernicus’ heliocentric theory over religious geocentrism, of Pasteur’s germ theory of disease over a belief in humors and spirits, of Darwin’s theory of evolution over the theory of divinely ordained fixity of species, and the triumph of modern molecular biology over vitalism. Three things have been true in each of these revolutions: 1. The phenomena being explained have been no less amazing and wonderful in the light of the scientific theory than they were in the light of the religious one. 2. The scientific theory empowered more accurate and detailed understanding of, and better control over, the phenomena. 3. The religious view was seen in retrospect as a kind of simplified approximation to the truth, still useful for some purposes, in much the same way as Newton’s mechanics approximate Einstein’s at low velocities.

I would like to suggest that the same may be true of theism: that it is a simplification, a commonsense approximation, of a truth that may one day be the subject of a scientific description. In particular, if, as the monists contend, the mind, or soul, is a computational process, and if, as the theists contend, the soul is of the same stuff as God, then it is a reasonable conclusion that God or the gods is or are computational processes as well.

Let me put this another way. A computer can be constructed out of many different kinds of materials: transistors, tubes, hydraulic valves, gears, billiard balls, and so on. It is difficult, in fact, to form a definition of the words “computer” or “computation” which rules out any physical processes at all, whether it be the motions of atoms in a gas or stars in a galaxy or the interactions of organisms in an ecosystem. Therefore, we are surrounded by computational processes, or mind-stuff, everywhere we look. Are some or all of these processes the mind of a god or gods? For example, the process of evolution on earth may be a computation which is the thought process of some great being.

So I repeat to the theists and dualists in the audience: don’t fear to let science play with your ideas. It may break them, but it is sure to return them to you repaired and in better shape than before.

Having delivered this speech from high in the lecture hall, I sat down next to Moravec, who was writing chapters for his book on a portable computer. The audience seemed briefly stunned, and then Eccles got up, smiling, and said he was glad to see that there were fanatics on the monist side as well as the dualist. I said I could live with that description.

Acknowledgment

The editors acknowledge the kind assistance of Mr Roy A. Varghese, editor of Truth magazine, in providing photographs of conference participants.

Note

1 For a copy of the proceedings or more information about this conference, contact the executive director: Roy Abraham Varghese, Box 59249, Dallas, TX 75229-1249