

ARTIFICIAL INTELLIGENCE with JOHN McCARTHY, Ph.D.

JEFFREY MISHLOVE, Ph.D.: Hello and welcome. I'm Jeffrey Mishlove. Our topic today is artificial intelligence, and we'll be looking at the past and at the future of this very exciting and yet somewhat arcane scientific discipline.

With me is Dr. John McCarthy, one of the founders of the discipline of artificial intelligence. Dr. McCarthy is one of the co-founders of the first artificial intelligence laboratory at MIT, and the founder of the artificial intelligence laboratory at Stanford University. He is the inventor of LISP, the major computer language used for artificial intelligence, and the oldest surviving computer language dealing with symbolic manipulation. He is also the individual who first conceived of interactive computer time sharing. He is the developer of non-monotonic reasoning, an important new form of logically conceiving of the difficult problems facing artificial intelligence today, and he is the 1988 recipient of the Kyoto Prize for his lifetime contributions to the field of computer science and artificial intelligence -- something of a Japanese equivalent to the Nobel Prize. Welcome, Dr. McCarthy.

JOHN McCARTHY, Ph.D.: Thank you.

MISHLOVE: It's a pleasure to be with you. Back in 1966 you wrote an article for Scientific American on the field of information, and you projected out at that time what we might see for the next twenty years, and, although there were some errors, I suppose, you rather accurately described many developments that we now take for granted, and which were at that time rather alien to the population at large. So you've witnessed a great deal of the history and the growth of a discipline which has dramatically touched probably everybody's life in the Western world today, and you've predicted that it would do so. I wonder if we can begin by just having you reflect a little bit on what these past twenty, thirty years have meant to you personally.

McCARTHY: Well, I've gotten older. I started my work in artificial intelligence in about '56, although I became really interested in it before that, in '49, when I was a beginning graduate student in mathematics. I would say that the field has made somewhat less progress than I hoped, although I didn't have any definite opinion as to how fast it would progress. I think that it had and still has difficult conceptual problems to solve before we can get computer programs that are as intelligent as humans.

MISHLOVE: One of the issues that you're working on, and to which most of your life has been devoted, is really tackling these problems - - providing the underpinnings so that we can ultimately have formal models of intelligence that would be equivalent to human intelligence.

McCARTHY: Well, that's right, and one part of the problem is to develop language in which we can express for our computer programs the facts and reasoning about the common-sense world that humans have, and that is necessary in order to behave intelligently. And I have worked on this using the tools of mathematical logic.

MISHLOVE: I think one of the striking things that I find in looking at the history of artificial intelligence is that in the early years there was some striking progress made on some rather difficult problems, like solving mathematical theorems. And people thought because we could do these difficult things, we ought to therefore have no trouble doing some of the simpler things that human beings can do. And yet just the opposite seems to have been the case -- that some of the simple things that any child can do, like recognize speech, have been the most difficult problems for computer intelligence.

McCARTHY: Well, the idea that one could really do difficult mathematical problems -- that is, creative mathematical things -- was not really realized; that is, it could do some simple kinds of theorem proving and things like that. Now, it's certainly true that dealing with

the common-sense world has proved to be quite difficult. What it amounts to is that while humans can do this kind of thing very readily because it's built into us, humans have much more difficulty understanding how it is done in order to be able to make computer programs do it.

MISHLOVE: You're developing formal models that deal with how human beings do the simplest of things.

McCARTHY: Well, there are two ways of looking at things. You can either look at it from the point of view of biology, or from the point of view of computer science. From the point of view of biology, you could try to imitate the nervous system insofar as you understood the nervous system, or you could try to imitate human psychology insofar as you understand human psychology. The computer-science way of looking at it says that we look at the world and we try to see what problems it presents in order to achieve goals and think about the world rather than about the biology per se. And I would say that the computer-science approach is the one that so far has had the most success, although these cannot be regarded as alternatives. They are like they're in a race, but they interact with one another; they help one another rather than hinder each other.

MISHLOVE: Well, in the field of psychology, it used to be during the fifties and early sixties that we thought of the mind as something like a black box -- you had a stimulus that went in and a response that went out. And I think it really wasn't until people in your discipline, artificial intelligence and computer science, began looking at how is this information processed that the psychologists themselves ever felt that they could have a handle on what cognition was all about.

McCARTHY: Yes, I think that's right. I think that Newell and Simon, who take an approach rather close to psychology, were the main contributors to getting psychology to move away from behaviorism. Behaviorism was a reaction to nineteenth-century philosophy, which really was very bad; but it went too far in its efforts to be scientific by

saying that the only things that were properly subjects for science were the things that were externally observable. But when computers came along, then it became clear that you couldn't do it that way. I remember there was an old computer called the IBM 704, and the only stimulus-response rule that it had was that if you pressed the start/read button a little yellow light went on. All the rest, to understand this computer you had to know what went on inside, and I guess computers have certainly had a profound effect on psychology.

MISHLOVE: As a psychologist myself, I'm very much a student of William James, who back at the turn of the century began writing about consciousness and the stream of consciousness, and I'm aware that for fifty, sixty years his work in that area was pretty much ignored, until people in the field of computer sciences began to say that we can have a handle on what consciousness means.

McCARTHY: Well, there are many kinds of consciousness. In some respects computers are easily more self-conscious than human beings. It's not hard to make a computer program look at its own program, but all that people have managed to do with it is to check some to see that it hasn't been damaged so far. What's involved in the kinds of consciousness that people would like to program is regarding the self as an object in the world, and to be able to think about what progress it's making toward achieving its goals, and so forth. And this offers some conceptual difficulties. I certainly wouldn't say that the problem of giving computers self-consciousness is very close to being fully solved.

MISHLOVE: Well, it raises some very deep philosophical issues. I'm aware of the disputes that took place during the 1940s, with Alan Turing, who was one of the founders of the field of computer science. He developed the famous Turing test, which suggested that if a computer could imitate a human being to such an extent that if you were sitting at a teletype, you couldn't know whether you were communicating with a computer or with a real human, then you

might as well say that the computer was in fact conscious. Turing responded, as I understand the argument, by pointing out the old philosophical conundrum of solipsism -- that we can't even be sure that another human being is conscious, let alone a computer.

McCARTHY: I don't remember Turing discussing solipsism, but he did use that as a kind of test for philosophers. In other words, if you wouldn't admit that something that you couldn't tell from a human was thinking, then maybe there wasn't much more to say. Now in fact up to this very day some of the philosophers are willing to accept behavioral criteria, and others are not; they even say, well, it could pretend to be a human, but it wouldn't really be thinking because it would only be doing what it was programmed to do.

MISHLOVE: One development -- and I must confess it troubles me a little bit -- that has come out of the information-processing models of the mind that are now current, is that we view consciousness as consisting of many components. You have memory, you have emotion, you have different kinds of attention, and sometimes people say, well, consciousness is nothing more than the sum of its parts, so to speak, just like a machine might be. I wonder how you respond to that.

McCARTHY: Well, a machine isn't the sum of its parts. If somebody took a car apart and gave you a heap of the parts, that wouldn't be a car. They have to be connected in a specified way and interacting in a specified way, and so if you want to say that the mind is a structure composed of parts interacting in a specialized way, I would agree with that, but it isn't just a heap of them.

MISHLOVE: It's more of a system.

McCARTHY: That's right.

MISHLOVE: Now we get into the issue -- and I know many people get offended when they think that you could even describe a human being as being equivalent to a system. People say we have something

more -- we have intuition, we have spirituality, we have something that transcends the mechanistic aspects of our being.

McCARTHY: Well, that view has been in retreat for several hundred years, as more and more was discovered about human physiology and psychology, and I suppose -- well, maybe one could use the boxing metaphor: it can run, but it can't hide.

MISHLOVE: Can you elaborate on that? I'm not quite sure what you're getting at there.

McCARTHY: Well, there are these aspects of human consciousness that have not been realized in machines, in computer programs, and there are some difficult problems for their realization, but we optimists about AI expect to get to them.

MISHLOVE: You know, there's an interesting story about you. You're a chess player, and back in 1968 you made a wager with a fellow who was then the Scottish chess master, that in ten years a computer would be able to beat him. And ten years later you got together with a state-of-the-art program, back in '78, and the computer nearly beat him.

McCARTHY: That's right. He won two games to the machine's one. Now, at that time David Levy was a graduate student in computer science, and my intention was not merely to bet with him, but to hire him to work on the chess program. However, he decided he'd really like to publish a magazine on chess rather than continue as a graduate student in computer science. I didn't consider the bet by any means a sure thing, but it came close in '78, and now -- this year, I believe, or maybe it was last year -- a computer program won its first game against a grand master, and since David Levy never made it to be grand master, probably the current programs could beat him, although in my opinion they use too much brute force in it.

MISHLOVE: Pure calculating power.

McCARTHY: That's right. I would like to see contests that are more like the one-design sailboat contests, where it's the cleverness that's involved, rather than who can build a monster special-purpose machine.

MISHLOVE: Now, this is very important, because your work seems to be saying that in natural human life we use a lot of mental shortcuts; we don't solve problems by using brute intellectual force. We somehow have rules of thumb that guide us. And you're attempting to develop formalized logic that would enable machines to be able to sort of work in that fashion.

McCARTHY: Yes, that's right, and indeed the collection of problems on which computer brute force can be applied is rather limited. Most of the problems of common-sense reasoning are problems where there really isn't that much opportunity to apply brute force, or at least nobody's really figured out how to do so. I would say the central problem of artificial intelligence involves how to express the knowledge about the world that is necessary for intelligent behavior, and I've pursued mathematical logic as the tool. This has had its ups and downs in popularity. Now is definitely an up period; it's quite popular, and part of the reason for that is that in the late 1970s several people independently, myself among them, discovered ways of formalizing what we call non-monotonic reasoning, which greatly extended the power of mathematical logic in the common-sense area.

MISHLOVE: Now, I know many of our viewers are going to have difficulty with a term like non-monotonic reasoning, and yet it may be crucial to our understanding of some of the developments that await us in the future, so could you expand on that?

McCARTHY: OK. You have to say what it's "non." Ordinary logic has the property that if you can draw a certain conclusion from some premises, then if you add more premises, you can still draw that conclusion. So the set of conclusions that you can draw only

increases when you increase the set of premises; they don't decrease. Now, human reasoning and what we will have to make computers do doesn't always have that property.

MISHLOVE: That's what you would call monotonic.

McCARTHY: That's right.

MISHLOVE: Could you give an example?

McCARTHY: Yes. Suppose I tell you that I have a bird that I want you to build me a birdcage for, and that's all I tell you. Then you would draw the conclusion that my bird can fly, and that you'd better put a top on the birdcage. On the other hand, if you learn the additional fact that my bird is a penguin, then you would feel that you do not need to put a top on it. So the conclusion that the birdcage required a top depends non-monotonically on the fact that I tell you.

MISHLOVE: In other words, this is an example of non-monotonic logic, and it has sort of built-in assumptions that I work with -- that is, when you use the word bird I assume it can fly.

McCARTHY: That's right, and that's the sort of convention of English or of other natural languages. If I hire you to build me a birdcage and you build it without a top and I refuse to pay, and you tell the judge, "He never said his bird can fly," the judge will side with me. On the other hand, if you did build it with a top, and I say, "Well, my bird is a penguin; he wasted material," the judge will side with you, because it's a convention of English that if a bird can fly, it doesn't have to be mentioned, even if it's important; whereas if a bird cannot fly, then it must be mentioned if it's relevant.

MISHLOVE: It reminds me of when I was a child, my father would sometimes say things to me. I would ask him a question, and he would say, "If you have to ask, if you don't already know, it won't help to tell you." We operate in a world of all sorts of implicit built-in

assumptions and built-in understandings, of context everywhere we go. And that is what I guess you mean by non-monotonic.

McCARTHY: Well, non-monotonicity is only part of the context problem. Now that we can formalize some non-monotonic reasoning, we see that, well, there's a good deal more to context than that, and that, so to speak, is the next mountain that has to be surmounted.

MISHLOVE: So you are attempting to use non-monotonic reasoning as a mathematical tool for building into computers an awareness that when I use a term like bird, I don't have to specify all of its qualities -- which ones they have, which ones they don't have -- in a straight, I would say linear, and you would say monotonic, specific form.

McCARTHY: That's right. Suppose you want to be able to reason about birds flying. Then you might say, "Well, I'll put in the exceptions -- ostriches and penguins, and so forth." And then someone comes along and says, "Well, what about a bird with its feet encased in cement?" And then you can see that you couldn't possibly put in all the exceptions, because if you put that one in, I'll invent another exotic exception that you wouldn't put in. So what you have to do instead is go to a system where you will assume that the bird can fly unless you have some evidence to the contrary.

MISHLOVE: Now, I have to tell you, this seems very simple to me, yet you're describing this as somehow something new in the world of computers.

McCARTHY: Well, that's right, and what I believe is that if it takes two hundred years to achieve artificial intelligence, and then finally there is a textbook that explains how it's done, the hardest part of that textbook to write will be the part that explains why people didn't think of it two hundred years ago, because we're really talking about how to make machines do things that are really on the surface of our minds. It's just that our ability to observe our own mental processes

is not very good and has not been very good. We can look at that historically, when we look at Leibniz, who was an extremely smart scientist; he was the co-inventor of calculus with Isaac Newton. He wanted to make a logical calculus that would permit calculation instead of argument, and he invented binary numbers in this case, but he didn't even invent propositional calculus; that was invented by Boole one hundred and fifty years later. And then Boole didn't invent predicate calculus. So what one sees is that each step in understanding of thought processes has taken time.

MISHLOVE: In fact what you're saying reminds me very much of the story of Socrates, who went around questioning people in all the different professions in Greece twenty-five hundred years ago, and discovered that while these people were quite competent at what they did, he said, "Well, they're all ignorant; none of them can tell me how they do what they do," when he questioned them closely. You seem to be engaged in a process very much like the second step beyond what he was doing.

McCARTHY: Yes, well, Socrates was as I understand it mainly interested in demonstrating people's ignorance, but now we are really trying to say, well, how can we make computers actually carry out these processes?

MISHLOVE: So it's a whole different program, in a sense.

McCARTHY: That's right, yes.

MISHLOVE: What is your sense of the likely future? You described twenty years ago how with computer time sharing we would all have access to information utilities, and that has come to pass. I don't want to pin you down to a specific date, because I realize you couldn't say honestly, but what are the sorts of things that are achievable?

McCARTHY: Well, as I say, I think there are conceptual breakthroughs that have to be made, and one extreme is that some smart young

fellow has just done it; he just hasn't told us yet. And the other extreme is that it may take a couple hundred years, maybe five hundred, even, depending on how many conceptual problems there are -- that is, it might take five hundred years before we have computer programs that are as intelligent as human beings. Now, I'd really be inclined to bet on something like fifty, although it's exceedingly unlikely that I'll be around. But I simply don't know how long it will take.

MISHLOVE: But it sounds like you're making a firm bet against the critics of artificial intelligence, who say that in theory it's philosophically impossible to replicate human intelligence.

McCARTHY: That's right. I see their arguments as faulty, and I don't see that human intelligence is something that humans can never understand.

MISHLOVE: So ultimately the project that you and your colleagues in the field of artificial intelligence are engaged in, one might view it as the most noble project of all, the one that Socrates actually urged people into, which is to know thyself, and sometimes against great odds to attempt to do what may be in effect one of the most difficult tasks facing humankind.

McCARTHY: Well, it's certainly a difficult task.

MISHLOVE: John McCarthy, we're just about out of time right now. I want to thank you very much for coming here and sharing yourself with us this evening. It's been a pleasure being with you.

McCARTHY: Well, thank you for inviting me.

MISHLOVE: Thank you.