An Interview with Oliver Selfridge[1]

Oliver Selfridge was born in 1926 in London. He studied Mathematics at MIT under Norbert Wiener and went on to write important early papers on pattern recognition and machine learning. His 1958 paper on the Pandemonium system is regarded as one of the classics of machine intelligence. He has worked at MIT Lincoln Laboratory, BBN and GTE Laboratories where he was a Chief Scientist. He has served on various advisory panels to the White House and numerous national committees. As well as his scientific writings, he has authored several books for children[2].

This is an edited transcript of an interview conducted on the 8th May 2006.

Phil Husbands: Could you start by saying a little about your early years? Were there any particular influences from home or school that put you on the road to a career in science and engineering?

Oliver Selfridge: Well, an important part of my education was my father. Without knowing any mathematics himself, he was wildly enthusiastic about my interest in it, which started at quite an early age: seven or eight. As was usual in England back then, I went away to school when I was ten. At the age of thirteen, I entered Malvern College, one of the (so-called) public schools. I remember we spent the year of 1940 in Blenheim Palace because the Royal Radar Establishment (RRE) had taken over the school. While at Malvern I covered calculus to the standard you'd reach after the first two years of a degree at MIT. One of the great things about education back then, and I am not sure that it's true any more, is that if you were good in one subject they'd move you ahead in that subject. You didn't have to worry about being good in both mathematics and French (which I was very bad at). So I'm very grateful to the English school system, although I didn't know it then; and I hated going away to school, of course, as I think everybody did. After Malvern I came to this country (USA) and started at MIT after a year and a half at Middlesex School in Concord, Mass.

PH: What year was this? You were quite young when you started at MIT, weren't you?

OS: This was 1942. I was just sixteen and the youngest in my class by more than a year. Last year we had a sixtieth reunion – the class of '45.

PH: What brought you to MIT? Did you go to the States because of the war in England?

OS: The Selfridges originally came from this country. My grandfather was born in Ripon, Wisconsin. He worked for a big store in Chicago called Marshall Fields and became executive vice president at an early age because he was, I guess, smart as hell.

---

[1] Preprint of P. Husbands (2008) An interview with Oliver Selfridge. In P. Husbands, O. Holland, M. Wheeler, (Eds), *The Mechanical Mind in History*, MIT Press, 397-408.

[2] Since this interview was first published Oliver Selfridge has died (3rd Dec. 2008).

He went on to own another store, which he sold, and then he moved to London where he opened Selfridges, a department store on Oxford Street. He borrowed a million pounds in 1906 or 1907, which was a lot of cash back then, and the store opened in 1909.

PH: And is still going strong.

OS: Still going strong, although there are no Selfridges in it! We lived in Kensington and then out in Norwood Green; there were four of us siblings. But then we came to this country because my father and grandfather were kicked off the board of directors of Selfridges at the end of the 1930s or thereabouts. My father came back to the States, because he had always been an American citizen; my grandfather had switched and become a British citizen in 1934. My father ended up working for a firm here called Sears Roebuck. Anyway, I went to MIT more or less by accident because I was very interested in mathematics and science. So I entered MIT at just 16 and graduated at 19, having specialised in mathematics. I went through the V12 programme which meant I joined the navy as a junior when I turned 17, or something like that, and they kept me at MIT, paying all the bills, which was wonderful, and then I went and got a commission in the navy just after Japan surrendered. After the navy I went back to MIT, to graduate school. I was working with Norbert Wiener and my friends Walter Pitts, Warren McCulloch and Jerry Lettvin were also there. By the way, I recommend the recent book *Dark Hero*[1] as a good source of information on Norbert Wiener. Anyway, by this time Walter had written the very important paper with Warren McCulloch, who was already a very well known neurophysiologist, showing how a neural net could do computations[2]. That came out in 1943 when Walter was only 19 or 20. I was very lucky to have met these people and then of course at graduate school I was introduced to a lot of others.

After that I joined Lincoln Lab, which was also a part of MIT, where we built the first spread spectrum system under Bill Davenport. Let me explain what that is. Communications theory, channel capacity and ideas like that, had just started; Shannon had written about them in 1948. The notion was that you needed a certain bandwidth to carry a certain amount of information. A spread spectrum system uses a much bigger bandwidth for that amount of information, and that helps to protect the signal, making it difficult to track or jam. We built the first system, which was classified, and the next ones weren't built for another twenty years. They are becoming more and more widely used now.

At that point, that would be 1953, I met Marvin Minsky who had just got through Princeton and was a Junior Fellow at Harvard. We were both very interested in what became known as Artificial Intelligence. He worked for me at Lincoln Lab for a couple of summers before he became a professor at MIT. Marvin and I ran the first meeting on Artificial Intelligence a year before the Dartmouth conference at the Western Joint Computer Conference[3]. At about this time, 1954 I think it was, I met a psychologist from Carnegie at The Rand Corporation in Santa Monica: Allen Newell. After talking for a couple of hours we had dinner that evening and he really appreciated what we were trying to do and he turned on fully to AI and started working on symbolic AI, which was different from what we'd been doing. Of course Allen, who died alas some time ago, became very well known, a very powerful guy. He was incredibly bright. Allen was terrific. He gave one of the papers at our 1955 meeting.

PH: I'd like to come back to Dartmouth and early AI later, but can we rewind slightly at this point to talk a bit about the origins of your celebrated Pandemonium system?

OS: I first presented that at the Teddington conference[4]. Do you know where the word comes from?

PH: I believe you took it from Milton's *Paradise Lost*.

OS: That's right. From the Greek for all the demons, it's mentioned in the first couple of pages of *Paradise Lost* which was written in 1667, I think. I wasn't alive then, it just sounds as if I were.

PH: The Teddington Mechanisation of Thought Processes Symposium was in 1958, but when did you start working on the system? Was it much before that?

OS: Well, we had been thinking about the general techniques of cognition for a while. The first AI paper I'd written was on pattern recognition, elementary pattern recognition and how to do it, and we spent a lot of time talking about it and getting people interested, and actually I work incredibly slowly. The cognition aspect was first sort of tackled by McCulloch and Pitts in their papers in 1943[2] and 1947[6]. So I talked with Walter a lot about certain things in cognition and the first paper on my work on pattern recognition systems was at the 1955 Western Joint Computer Conference[5]. So Pandemonium incorporated many of the ideas I'd been developing. It's an idea that is very powerful and people like it, but nobody uses it.

PH: It's a really impressive piece of work; the paper pulled together a lot of very important ideas in a coherent way – parallel distributed processing, adaptive multi-layered networks, feature detectors and so on. I'm curious about the influences, the currents that came together in that paper. For instance, you knew Jerry Lettvin very well and during that same period he was working with Maturana, McCulloch and Pitts on the research that produced the landmark paper *What the Frog's Eye Tells The Frog's Brain*[7] which gave a detailed functional account of part of the Frog's visual system and demonstrated the existence of various kinds of visual feature detectors suggestive of 'bug detectors'. It seems to me there are quite a lot of connections between that work and Pandemonium. Is that right?

OS: Oh absolutely. In fact if you look at their paper there is an acknowledgement to me, and I acknowledge Jerry in the Pandemonium paper. They were influenced by my pattern recognition work and the ideas behind it, which were to do with cognition. We regularly discussed the work. The question is about cognition — what does the frog do when he sees. Many people still think that the retina merely detects pixels and ships them off to the brain, which of course is just not true. The frog's eye paper was published in the Proceedings of the IRE, now the IEEE, because the Journal of Neurophysiology wouldn't accept it; they said it didn't have real data in it, like numbers. Well of course it didn't, it was much better than that. I remember we laughed about it. Jerry built the first microelectrode needles for reading from single axons in the frog's optic nerve. It was an absolutely brilliant piece of work both in terms of the ideas and the experimental manipulations.

Of course Jerry and I were roommates while I was in graduate school, along with Walter Pitts. It was always exciting.

PH: That was quite a combination.

OS: Well I had a good time indeed. Walter and I often went places together. One summer, I think it was `48, we climbed the Tetons in Wyoming just before spending the rest of the summer with Norbert in Mexico city.

PH: The frog's eye paper is often quoted as containing the first full statement about low-level feature detectors in a vision system – moving edge detectors, convexity detectors and so on – building on Barlow's earlier work giving evidence for 'fly detectors'[8]. This notion became very important in vision science. Did you play a part in that, since you were using the idea of feature detectors in your pattern recognition systems?

OS: Well in some sense I probably did, but a lot of other people came to it independently. My first paper on pattern recognition included the question of how you recognise a square. It described how the features of a square include a corner and a line and asks how do you detect a line against a noisy background and so on. So yes I was the first one to put it in specific enough terms that it could be computerised, as far as I know, but I think a lot of others came up with it independently … of course this was fifty-three, fifty-four years ago, so not quite BC, but getting that way.

PH: So maybe the idea was floating around to some extent, but it seems you made a very important contribution and obviously influenced the Lettvin/Maturana work.

OS: Thank you. Well, Jerry and I have always been on very good terms and I knew Maturana quite well, but he went off and had an independent life of some notoriety. Walter Pitts, of course, fell apart. That was tragic, really tragic. I'll tell you the story very briefly. In 1952 Norbert Wiener accused us—Warren McCulloch, Walter and me—of corrupting his daughter, Barbara Wiener, who was a year younger than I, based on what Norbert's wife, Margaret, told him. She didn't like us because she thought we were too free and so on. The accusation was absolutely false. Norbert then turned against us and wouldn't speak to us or acknowledge our existence for the rest of his life, which was a great tragedy. Now Walter fell to pieces because of that, because he was dependent on Wiener. Walter had the highest IQ of anyone I've ever met, but he was fragile. When Walter was about eighteen or nineteen he bumped into Norbert Wiener and greatly impressed him with his mathematical ability — he corrected something Norbert showed him — and so he started working with Wiener and they became very close. Anyway, you can read more about their relationship in *Dark Hero*. Then after Norbert wouldn't speak to us, I remember being at a party somewhere in Cambridge with Walter and he said "I wonder why people smoke. I'd better try." Two weeks later he was two packs a day. So he sort of fell apart and he played with drugs of all kinds and fifteen or so years later he died, essentially of overdoses. He was a total genius but he didn't know how to handle himself at all in a social way. It was just terrible.

PH: Pitts is reported to have destroyed most of his work from that time, so many of his ideas never saw the light of day. Is that true or did some of his work live on through his influence on people like you who worked with him?

OS: Well, it's pretty much true. But he did a lot of other interesting things. For instance, there was at MIT a professor called Giorgio De Santillana, a historian and philosopher of science, whom Walter spent a lot of time working with later when he had his personal problems. His inspiration for Jerry was quite real. The full list of authors on the frog's eye paper is Lettvin, Maturana, McCulloch and Pitts. The work was done in `56/`57 and he still had a real input at that time. Incidentally, something that pissed everyone off, including me, was that Hubel and Wiesel took the genius of the ideas and the genius of the microelectrodes and the experimental setup, and they got a Nobel prize. In their Nobel prize speeches they did not give any credit to Jerry. That was rotten manners, putting it very mildly.

PH: During that period, in the 1940s and 1950s, you interacted a lot with at least two people who have had very important influences in neuroscience: Lettvin and McCulloch. Was this more by accident than design or did you deliberately work in an interdisciplinary way?

OS: Sort of both. The number of people interested in these things in the mid 1950s wasn't very large, and so we tended to know each other and talk to each other. Norbert and Warren and others had initiated interdisciplinary ways of thinking and that was still around. AI had only just started at this point and new people, such as John McCarthy, were coming in. Claude Shannon was still interested, although he soon stopped. Von Neumann was interested, although he'd written all his papers by this time. He became a devout Roman Catholic in 1955 when he was suffering from cancer. Warren McCulloch kept going although his papers got less specific and I think less useful, too general. By the late 50s he was drinking a quart of scotch a day and you can't do that and keep your mind working as well; at least he couldn't. Maybe it's a good way to go.

PH: You seem to be making a clear distinction between AI and Cybernetics. Is that how you see it?

OS: Yes, very much. Cybernetics obviously preceded AI; in fact Jerry Lettvin and I are probably the only two people left alive who are specifically mentioned in Wiener's *Cybernetics*[9]. The notions of Cybernetics are in AI but the focus is different. Cybernetics turned out to be much more an engineering business than AI. There is a great deal of engineering in AI and all the major thrusts that we now have are based on mathematics, but that is not what AI is about. A lot of AI you will see expressed in mathematical terms, but many of those aspects pretty much ignore what to me is the key power of AI, which is learning. Learning is central to intelligence.

PH: In a nutshell, how would you define AI?

OS: I think it's about trying to get computers, or pieces of software, to exhibit the intellectual powers of a person. That's a vast range of things, so to me the deep key is learning. It comes out in three very different aspects: the actual actions you take, the cognition, and the memories of experience. There is a special action part of experience which is planning. In an intellectual sense planning is done only by people. A key thing that we are working on now is the essence of control as part of action. How do we learn how to control things? I think the essence of control is purpose — you want to do something. It's not just that you like beauty or you like good art or something, it's that you have a whole structure of purposes. If you're right-handed and you hurt your right hand so you have to use your left, you can still pick up a cup of coffee without thinking about it. The purpose is to get coffee to your mouth. This means you have sub-purposes of finding where the cup is, moving your arm and so on and so forth: it's purposes all the way down and also all the way up. But those purposes change all the time and the essence of control is trying something and improving it. As Marvin Minsky said, “The best is the enemy of the good”. Because ‘the best’ implies a static universe, but it ain’t static. The problem with a lot of the mathematical treatments is that generally they are looking for formalistic presentations of processes that can then be optimized. But we don’t optimize, we improve. To me that should be part of the essence of AI.

But AI, like any other science, is a very complicated thing. In physics, Newtonian mechanics is a perfectly adequate way of expressing many processes, such as shooting a gun or something like that. But it turns out that in a deep sense Newtonian mechanics is just wrong. But looked at another way it isn’t wrong exactly; for certain purposes it gets improved. The same thing is true with AI. For certain purposes the simple memories we have about what we did, and why we wanted to do it, are adequate. But often, next time we do something we are trying to do it differently, or we modify it. It’s very hard to think of something that we don’t do better the second time. Likewise it’s very hard to think of a computer task that the computer *does* do better the second time. So I think that in AI we should work on developing software which will notice what it does, remember the experiences and what it wants during the experiences, and be able to improve. Not just the actions, but the cognition and the planning too.

PH: You put learning and adaptation at the heart of AI, so looking back over the past fifty years do you think the trajectory of the field has been reasonably sensible or do you think there have been some disastrous directions?

OS: No, not disastrous. We’ve done a lot of powerful things but they’re missing a great deal of what I’m interested in now: purpose. You raise your children by encouraging and motivating them, but how do you encourage a computer programme? To use a high tech Americanism, the programme doesn’t give a shit. Well, your children all did and still do. So that is what I’m working on now and what I think is important. Marvin’s *Society of Mind*[10] discussed some of these issues, in very different terms, twenty odd years ago. But I’m trying to be more specific and we’ll see if I live long enough to get these ideas in any kind of shape.

Learning and adaptation have certainly been constant themes throughout my work. Adaptation I regard as a special case of learning, an affirmative case. For instance, the motor cortex makes a muscle move without affecting it directly—there is a loop out from the spinal cord to, say, a finger muscle with the signal coming back to the spine, so that we have a control circuit. The motor cortex modifies the gain of that circuit so it’s adaptation all the way down, so to speak. We don’t necessarily need to go as far as that; indeed I think copying all the details of neurophysiology is a silly error, but understanding what happens and why is the thing. Most people in AI don’t do that. When I give a talk many people agree with me but then they go back and do the old things. Most computer programmes are full of errors with no way to correct them. Well I want a piece of software that can limit its errors by learning, and thereby try to correct them.

PH: During the Cybernetic period and in the early days of AI there was a lot of interest in adaptive and learning systems but that seems to have greatly diminished by the late 60s and the pattern continued throughout the 70s. Why was that?

OS: It was regarded as too hard. When Feigenbaum and Lederberg developed expert systems in the late sixties there was almost no learning involved. The learning was confined to the people. I have a very high regard for what they did and don't object to it, but my feeling was and is that learning is the key, and a lot of the deep questions were ignored. But work like that did bring a lot of people into AI, and I want more of those people to turn to basic research questions again. As I've said, I think purpose and motivation are the deepest requirement that we need in AI now: you want the software to care. People might say, "Well my system has the goal of winning as many games as possible, isn't that caring?" Well, yes, sort of, but it's only the beginning, why do we stop there? We still can't really usefully praise or reward a system.

PH: Looking ahead and speculating, do you think the sorts of architectures and methods used in AI today will have to be abandoned or radically changed to make significant progress?

OS: Well I think we will get to the point where AI has some sort of reward structure that enables it to learn in a more sophisticated way and then we won't so much programme our systems as educate them. That will work and it will work spectacularly well. Communication will be very important as pieces of software will also teach each other. Getting motivation and caring and being able to adapt on multiple levels will be big breakthroughs, but it will require more than that - there won't be just one thing. But we need to get started. It will also have to make money for someone because funding for pure basic research is very hard to come by today, certainly in this country.

PH: Do you think AI will need to get closer to biology to make these advances, or maybe move further away?

OS: Well I don't think we need to move further away. There is a big effort now in neurobiology and computational methods are playing a part in that. A lot of the effort is looking at single neurons in detail. I'm not sure that will help us get AI. There are too many steps from understanding a single neuron to having intelligence. That isn't to say that we can't learn some very important lessons and take very useful ideas from understanding more about how the brain works – just as happened, for instance, with Jerry Lettvin's work— but I think it has to be at a higher level than single neurons. Of course the picture keeps changing in neuroscience anyway. The recent discovery of the important functional role of glial cells is an example; in essence they really have to start thinking all over again and come up with a new explanation.

PH: So you think detailed modelling is too ambitious but taking inspiration at a more abstract level is useful?

OS: Yes. Detailed modelling is too ambitious and won't work. But more abstract inspiration is very important. Absolutely. Two important biologically inspired areas are of course neural nets and John Holland's genetic algorithms. There is a lot of stuff going on in both areas and a lot of it is very successful at solving problems, but there are great limitations and simplifications in these areas as they stand today. One fault is the emphasis on a single evaluation function. You need multiple purposes at different levels and multiple ways of evaluating these at different levels. It's time to try and tackle issues like that.

PH: Related to this, how would you say your interests are divided between developing artificial intelligence and understanding natural intelligence?

OS: Oh equally. I'm interested in both, that's always been the case.

PH: This year is the fiftieth anniversary of the Dartmouth conference and there is a lot of talk again about it being the birthplace of AI and all that. From what you've already said here and elsewhere, that's obviously an over-simplification as the basic ideas were already around or being developed, you had your West Coast meeting in 1955, the name AI had already being used by some of you and so on. So do you think anything much actually came out of Dartmouth itself or was it more a part of an ongoing process?

OS: Both. Dartmouth generated a spectacular amount of interest because it got a lot of publicity. People were persuaded to look at new problems and Allen Newell convinced a lot of people that symbolic processing and reasoning was important. So it was a very effective step; it got national interest, much more than Marvin and I had got for our earlier meeting, and it spread the message around. There were a lot of interesting and powerful people there: John McCarthy was a founding trigger of the meeting; there was Nat Rochester from IBM, and many others.

PH: Presumably the publicity and interest were helpful in generating funding.

OS: Well funding didn't follow particularly speedily, but yes Dartmouth did help in that respect, it opened various people's minds to the possibilities.

PH: Finally, is there any particular piece of work of the many that you have been involved in that stands out for you.

OS: Well not exactly, but I suppose the Pandemonium work is special to me because it helped me to finally nail a lot of issues.

[1] Jim Siegelman and Flo Conway (2004), *Dark Hero of the Information Age: In Search Of Norbert Wiener--Father of Cybernetics*, Basic Books.

[2] McCulloch, W.S. and Pitts,W. (1943) A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, **5**:115—133.

[3] The Western Joint Computer Conference, Los Angeles, 1955.

[4] Selfridge, O.G. (1959), Pandemonium: A paradigm for learning. In Blake, D., Uttley, A., (Eds),*The Mechanisation of Thought Processes. Volume 10 of National Physical Laboratory Symposia*. Her Majesty's Stationary Office, London, 511-529. [Proceedings of the Symposium held at the National Physical Laboratory, Teddington in 1958].

[5] Selfridge, O.G. (1955), Pattern recognition in modern computers, In *Proceedings of the Western Joint Computer Conference,* ACM:New York.

[6] Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, **9**:127-147.

[7] Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959) What the frog's eye tells the frog's brain, *Procedings of the I.R.E.* 4**7**: 1940--1959.

[8] Barlow, H. B. (1953) Summation and inhibition in the frog's retina. *Journal of Physiology*, **119**:69-88.

[9] Norbert Wiener (1948). *Cybernetics, or control and communication in the Animal and the Machine.* MIT Press.

[10] Marvin Minsky (1986) *The Society of Mind*, Simon and Schuster, New York.