Invited Speaker: Bill Clancey

by Jacobijn Sandberg

On October 18th Bill Clancey gave a talk at the DELTA-conference held in The Hague (The Netherlands). His talk mainly deals with the notion of 'situated cognition' and its implications for Artificial Intelligence and learning. Because the ideas surrounding situated cognition give rise to a lot of debate we hereby present an excerpt his talk and we explicitly invite our readers to react to the ideas presented. Of course the presentation Clancey gave was much more elaborate than can be represented here.

"Imagine that you would like to develop a program that can draw pictures.

[A picture is shown of three people in a garden]

This particular picture was drawn by the AARON-program developed by Harold Cohen. Imagine the difficulty that Harold Cohen has. He would like to create a kind of robot that can move a pen around on a piece of paper or on a screen, and he would like the robot to create original drawings.

The problem that Harold has is that, if the pictures have to be original, then we cannot put into the program descriptions of what the pictures will look like. So we can't plan for the program in detail. The program has to be the original creator of the representation. We can't put a grammar inside either, which would generate certain kinds of pictures, because again we would be predescribing the program's behavior. In essence, Harold Cohen was face to face with the ultimate limitations of the Artificial Intelligence approach. The vast majority of the programs that we see today are based on a human designer putting a kind of grammar into the program to specify what the program will be capable of doing. So Harold had to rethink what it meant to create a representation, which of course we all knew in Artificial Intelligence was something we hadn't resolved. The essence of his approach was to realize there is a difference between what the observer would see when the picture was complete and the attitude or the orientation of AARON as it was drawing the pictures. Indeed, the pictures don't exist until they are on paper.

I will give a number of examples in my talk to illustrate this change in orientation. It makes us realize that representations are not quite what we thought they were. What we have put inside the head of the agent must really be produced by the agent in his or her behavior. We will see in my presentation that we are reconsidering what knowledge is, what knowledge representations are, and therefore what learning is. In general, most of the statements and drawings that we have said are an indication of knowledge are created every time they are used freshly; memory is different from what we thought it was.

Well, this is quite a mouthful, and you will see that there are many angles here. I am talking about a different view of knowledge, representations, even information, a separation of what is going on in the mechanism and what we see, and this has very exciting and interesting implications for learning technology. It has implications at the biological side, if we want to build robots or understand human perception, and it has implications at the sociological side, because we see that the environment plays a very different role than what we expected.

I am at the Institute for Research on Learning, which was formed with the idea, in essence, of bringing together the social and cognitive models of human reasoning and behavior. We have social-anthropologists, social-linguists, educational psychologists and me, the AI researcher. The problem is for us to learn how to talk together and to collaborate on different projects. I like to call this field situated cognition. Perhaps its strongest manifestation today in AI research is called situated robotics or situated automata

A lot of the work originated in linguistics and anthropology back in the 1930s, the work of Bartlett, George Mead; in philosophy, the work of Wittgenstein and Heidegger is very important. In sociology there is the work of Durkheim and Mannheim, and of course Marx. Situated cognition research in the workplace in the United States has come mostly from Scandinavia. The ideas of Scandinavian design, how to use technology in a social context, have been very influential for the people who formed the Institute for Research on Learning.

Here is the outline of my talk. I will start by saying something about Artificial Intelligence, because we are not just going to throw away the old ways of building programs and the old ways of thinking. Instead, I believe we can generalize what AI-programming is in terms of a modelling technology. So we are going to reinterpret the models that have been generated and we are going to use that technology as another tool, just as we use, say, differential calculus or other forms of numerical analysis. We need to separate out AI modelling technology from the models themselves. Then I'll tell you more about the memory-as-stored-structures hypothesis which has driven cognitive science and our Intelligence Tutoring Systems (ITS) work in the last twenty years. I will argue for a kind of Copernican shift, as we see memory in a different way and we put representations out were they can be perceived, say as silent speech and visualisations, but mostly in our writing and in our drawing. I will give you some more examples of representations and then talk about the implications for expert systems and ITS research.

My first claim is that AI programming is a technique for modelling processes qualitatively; you can call it qualitative modelling or relational modelling. It identifies objects, events and describes them according to spatial, temporal and causal relationships. Whenever you look at an AI program you should be able to answer some questions about models.

First, we identify some system that is being modelled: biological system, manufacturing process, electro-mechanical process, the interface of a computer system, social organisation, and so on.

Then we can ask, what is the modelling technique? As I said, it won't be just a series of equations with a set of numerical parameters. Instead we will find some kind of relational network--most common is a classification model--which may describe stages in a process or interactions that the system has with its outside environment. This is what a disease-hierarchy is. Also very common are simulation models, often (too narrowly) called model-based reasoning. This kind of model generally takes the form of causal-transition networks or structure-function networks. This is really the computer science contribution of AI-programming, this new way of representing processes in computer programs.

Continuing, we can ask why is the model constructed? It could be for designing some new system, for assembling that new system or, given an existing system, diagnosing why it is not behaving the way that it is supposed to, or for predicting and controlling that behavior. One of the things we now realize is that beyond traditional numeric modelling--which is very good for prediction and control--we are now able to add other tasks, other things we can do with a model, using qualitative modelling.

Next, there are different techniques for constructing these models; heuristic classification is one method I have distinguished. Finally, we want to consider the context in which the model will be used and constructed. This is part of the new thinking about designing new programs. What is the context in which it will be used? What is the environment, what is the larger system in which this is just part of the solution? Some of the things we have to consider are how reliable is the data and its availability; this affects the modelling techniques that we will use.

To recap, if we look at an AI-program we'll find inside a model of some system in the world. Let's take for example a program for teaching medical diagnosis. It would have a model of pathophysiological processes. In my own work this was a classification model of meningitis and related diseases. Next, we have a model of the process of diagnosis (the task). This is a description, another qualitative model; and the diagnostic process description interprets the pathophysiological process description. This is one of the key ideas in AI-programming: The representation allows one process to inspect and reason about another process--perhaps set it up as a simulation and predict what it will do.

Here is an example of what we have been able to do with this modeling capability. This was actually the motivation for my research in the Neomycin program in 1980-1990. Suppose we have a series of questions asked by a student who is carrying on a medical diagnosis, and we want to know what is the knowledge that the student is using: What is his model of the patient, how does he do diagnosis, what conclusions is he making as he precedes through the series of questions?

As I claimed at the beginning, the technique here is grammatical. What we can produce from our descriptions of the diagnostic process is a kind of a parse, the familiar "deep structure" from natural language generation, or, more colloquially, a diagram of a sentence. At the terminal nodes we have the student's request for information: Did the patient have an X-ray, does the patient have rashes, did the patient undergo splenectomy, and so on. We parse that and say that the student was pursuing three hypotheses. There is a logic as to why these hypotheses were pursued in this order, which follows from finding out about the chest X-ray.

So this is my claim about AI programming technology--it's a way of modeling processes qualitatively. The dramatic shift is to say that this parse is our model, as observers, as we look at patterns in the student's behavior. To say that the student is actually creating these networks and carrying out these particular calculations is usually not true. However, all of our work in expert systems, knowledge acquisition work, and student modelling has been based on this hypothesis, that the structures that we develop, these models of reasoning, these models of processes in the world, are encoded in the memory of the human being and that they generate the behavior that we see.

The perspective I want to emphasize is that a knowledge representation is just that. It is a representation of knowledge. Yes, you can represent knowledge but you can never have knowledge in hand. Knowledge-level descriptions are capabilities that we ascribe to someone in order to explain why they behave the way they do. This is the definition used by Allen Newell in his paper "The knowledge level" in 1981. Or, as was said by a russian back in the 1940s: The map is not the territory.

But in AI research we look at the structures of our models and we say, "This is the knowledge; this program is an expert; this is what the student knows. It is these structures." In so doing, we have claimed an isomorphism. We have said that what is in the student's head and these representations are functionally identical. But if people literally followed such grammatical patterns or shuffled them about grammatically the way our learning programs do, they would not be very intelligent. We have confused our representations with the phenomenon we are modeling. The map is not the territory.

As I started getting into situated cognition research, something started sounding familiar to me. It was the physics of Heisenberg -- the uncertainty principle. In short, the world--reality--is created in interactions. All we ever have in hand are models of these interactions. It is through interactions that properties are created.

Well this is a controversial point in physics, and it is probably no surprise that if we go through anthropology, sociology, physics, linguistics and now AI we are going to find this same tension between objective and subjective points of view. What are models, and what relation do they have to the phenomena they are about? Are they reality itself? It should be no surprise, given what I have been saying, that I have a relational perspective, and it is the Heisenberg/Bohr interpretation of quantum mechanics.

There is something out there, yes, something apart from us as individuals. But it has no inherent properties, no objective substance or form. The form, the properties, are only created by the interactions objects (or, better, processes) have with each other. So, we build things like linear accelerators, cameras, thermometers, or video-cassette recorders and we can play things forwards and backwards. We create information by how we interact with our environment. We segment the world, breaking it apart, giving names to the patterns that we see. Of course, our own perception is biased by the categories in which we are already thinking. We write down descriptions of the processes that we observe, descriptions of the interaction. We are always describing patterns of one system interacting with another.

This is the first order of relativity. The correspondence claim, which is the common sense claim, is that the representations are identical to the reality. They describe in a one-to-one way the properties of the world as it is. Science, as most of you know, makes the same correspondence claim. It is in fact essential for us as scientists to have this kind of belief. We have to believe that there is something apart from us, that is outside, that we are getting a closer and closer description of. In fact this is the view of knowledge that is in textbooks: It is apart from us, it is not subjective, and it is the world that we want students to know about. I am saying quite the contrary: We created the properties by our instrumentation and we created the patterns in our perception. There is even a secondary form of relativity, which is that a reader will look at the text and interpret it, also with biased perception. Discarding the correspondence view of reality, the view of models as being reality, of knowledge as consisting only of models, is essential if you are to understand situated cognition.

Let's consider a concrete example of this from knowledge engineering. When we have a set of representations in our program, what we consider to be a case, the data, is determined by the model that we have. The model indicates what the relevant observations should be. Now, at a certain point the cases may cause a breakdown, so the new cases will cause our old model to fail. So then we go back and perhaps generalize our representations. Or we might have to create new representations.

And now an important thing happens. When you put new representations in the program, you go back and look at the old cases and you need new data, because you have learned that new measurements need to be taken. One of my favorite examples is Legionnaire's disease. Until physicians knew about it they didn't know to ask, "Have you been to a convention where there is air conditioning?" So what constitutes information is determined by the model you have. This gives us a very different view of teaching a student a knowledge representation. What kinds of transformations might a student be going through as he is relating examples, previous cases, old representations, and new representations? How do cases give representations meaning? Representations aren't merely applied; they are created by acts of interpretation in every new situation, as perceptions, ways of seeing.

I am going to shift now to the sociological perspective. Another aspect of situated cognition research has to do with the tight coupling between perception and action. Our behavior is always at some level immediate. There are no intervening representations. It is when we cannot behave automatically that we start creating representations and these representations then orient what we see, what we pay attention to.

Here is an example of how this idea gets applied. The social scientists kept telling me that knowledge is in the environment; knowledge is not just in the head. So I tried to find a good example from my own experience.

Well, what am I good at, what can I do? I know how to go backpacking, camping, and I am very good at packing; I never forget things. When I go backpacking I take this box out of my closet, because this is were I have put things from the last trip, and I know these are the things that I should bring with me. I do prepare a list of things to bring, and I do think ahead a little bit, but mostly I go to this box and start reacting. I start seeing things; it reminds me of things. I might create a list as I go along, but my behavior is so data-directed, it is so much influenced by the materials that surround me, that if you want to predict what I am going to do, if you want to know the order in which I am going to pack, you need to study the structure of the box.

This is a very important perspective, which the social anthropologists have been pushing. I do write things down, and after every trip I make some notes about what happened at the last trip. But taking into account that box in my closet suggests a different perspective on memory and how we actually do things every day. Let us ask again, what does it mean to be an expert, to be able to perform? Is it a just matter of representations stored in the head? We must take into account how people structure their environment, for they know very well that this will bias their subsequent behavior.

Rather than talking about representations that we have never seen, that are hidden, stored in the mind, manipulated in ways that nobody has ever seen, we should at least pay some attention to the concrete things that people do to represent and structure their future behavior, things that are around people all the time: their notebooks, how they arrange their desk, piles they make that segment and prioritize their activities, and so on. This stuff is out where you can see it. Put simply, situated cognition research claims that behavior at the lowest level is purely reactive (as opposed to being mediated by representations). As Wittgenstein told us, even when we use representations, we must be able to do so without having to look up what they mean, or we would be caught in an infinite regress

Let's move on to consider implications for education. We want the student to be able to participate in some community of practice, to be a member of that community, to be able to work with them as they create their own models. To show you the implications let me now go back to the medical diagnosis example. From 1977 through the mid-eighties I said I was trying to teach a knowledge base to students. I said that the knowledge base is what makes you an expert; if you just memorize it, then you will be an expert.

Now look at how we think about our models in a different way. Here are things that I might want a student to be thinking as he sees a theory that someone gives about the world in which he is going to behave. Consider the disease taxonomy which was in the Neomycin program. The student might ask, well what are taxonomies good for; why do I need them? If I had a blueprint of the human body would I be able to throw away this taxonomy? That turns out to be false. What do taxonomies leave out? Is this simplified in some way; is this biased for a particular population of people in some region of the world? Who else knows this? Can I talk to a nurse about this, or will only a specialist know this particular thing; does a general internal physician know it? Do people agree? Am I going to find people as I travel to different parts of the world who have different views? And how do they find out that their theories are different; how do they resolve those differences? Is this what medical research is doing, so if I go into medical research, I could get involved in this kind of model building? And what should I be reading? Whom should I be talking to as these models change over time, as new equipment becomes available, and so on? To summarize, the view which we have been emphasizing at the Institute of the Learning Research is that learning is becoming a member of a community of practice. As these questions indicate, this involves much more than memorizing somebody's model of the world.

How does this change my idea of what an expert system is, or to put it in another way how should I use computers? Most AI researchers start by trying to automate everything. So when we see experts talking to less experienced people, giving advice, how do we use the computer? To replace the expert, of course. Most of us have moved pretty far from this, but this was the original idea of the Mycin consultation program and how expert systems got started. We use a computer to automate, that is what computers are good for.

Now a different view is a little more complicated. Imagine that you are building an expert system, say for the Digital Equipment Company, and you are John McDermott in 1981. Your first idea is to create a program which replaces the experts; it lays out computer systems or it checks the configuration of a computer system. The new perspective is to keep in mind that people will continue to talk to each other. Think of the program and the computer as something to put on the side, next to the people, which will mediate and facilitate their conversation.

So, here we have a sales person talking to a customer. They are designing something collaboratively that neither of them could design alone. Furthermore, they are not technical people; they do not know all the details of how the computer systems work. The computer program becomes a kind of electronic blackboard where they can post their model of the customer's working situation: the floors, the sites, what kinds of communication is necessary, the working groups, the kinds of information processing they do, and post the model of the DEC computer system, or whatever company, which is going to be sold to the customer. Next, they can look at this model and the customer can say, "Oh, I can't have three printers on one floor here. The person on the first floor needs to have output and she uses a word processor for this reason or another." So the sales person goes back and edits the program's output. He has just learned something more about the customer's world.

Actually we have modelling of two different communities going on. A second dialogue occurs when the sales person returns to the experts inside the company and says, "I keep coming up with this situation over and over again, and you don't have a product which people are ready to use." They can look back over the models that have been constructed, review the cases, and the representations and create new models. This kind of negotiation can occur for diagnosis, repairing, controlling, and so on. The representations provide a medium for the conversations between people.

It is not just one student, one computer, one user, but people naturally working together. Analogously, a different view of Intelligent Tutoring Systems would be to put the program between two students. We could have the program learn something about the students, and tell them to go off and talk to someone else, as it develops a model and sees patterns in what the students know related to each other. This is a very powerful model because people create models in conversation. To facilitate learning, we ask, how can we facilitate conversation and interactions between people? This is pretty far from the old idea of individualized instruction.

In conclusion, situated cognition is not coming from any particular area, any particular domain, or any part of the world. It is a synthesis of ideas that have been around for at least 70 years. I have shown you how it goes back to linguistics, anthropology, early work in social psychology by George Mead and Bartlett and, amazingly enough, work in physics and, perhaps not so amazing, work in philosophy. The linguists are telling us to see conversations in a different way; the anthropologists are telling us to look at the representations that people actually create and see where they are stored out in the world. Look closely at how representations change in conversational interactions. Donald Schon studied an architecture student and looked at the drawings and how they changed over time. Philosophers give us a different perspective on how to think about models and reality, which is very important for us as we use this AI modelling technology. Instead of thinking about the theory that we create as something that we deliver to the student and have the student actually memorize, the theory (a knowledge base) can be perhaps a specification, a process description of the kind of coupling that we would like to bring about between people. A medical diagnosis strategy can be viewed as a useful way of talking about interactions between physicians, patients, and technicians. But more fundamentally, we must reconsider how such representations emerge in social interactions. It leads us to a different way of thinking about technology and of putting it in the workplace."

References

Agre, P. E. 1990. Book Review of Plans and Situated: Actions: The Problem of Human-Machine Communication. Artificial Intelligence, .43(3), 369-384.

Bartlett, F. C. [1932] 1977. Remembering-A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press. Reprint.

Bateson, G. 1972. Steps to an Ecology of Mind. New York: Ballentine Books.

Bateson, G.1988. Mind and Nature: A necessary unity. New York: Bantam.

Bransford, J.D., McCarrell, N.S., Franks, J.J., and Nitsch, K.E. 1977. Toward unexplaining memory. In R.E. Shaw and J.D.

Bransford (editors), Perceiving, Acting, and Knowing: Toward an Ecological Psychology. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pps. 431-466.

Clancey, W. J. 1989. The knowledge level reinterpreted: Modeling how systems interact. Machine Learning 4: 287-293.

Clancey, W. J. 1991a. The frame of reference problem in the design of intelligent machines. In K. vanLehn and A. Newell (editors), Architectures for Intelligence:

The Twenty Second Carnegie Symposium on Cognition, Hillsdale: Lawrence Erlbaum Associates.

Clancey, W.J. 1991b. Why today's computers don't learn the way people do. In P. Flasch and R. Meersman (editors), Future Directions in Artificial Intelligence. Amsterdam: Elsevier.

Clancey, W.J. (In press.) Model construction operators. To appear in Artificial Intelligence.

Clancey, W.J. (In press.) Book review of Rosenfield's The Invention of Memory. To appear in Artificial Intelligence.

Cohen, H. 1988. How to draw three people in a botanical garden. Proceedings of the Seventh National Conference on Artificial Intelligence. Minneapolis-St. Paul, pps. 846-855.

Collingwood, R. G. 1938. The Principles of Art, London: Oxford University Press.

Iran-Nejad, A. 1987. The schema: A long-term memory structure or a transient functional pattern. In R. J. Tierney, Anders, P.L., and J.N. Mitchell (editors), Understanding Readers' Understanding: Theory and Practice, (Hillsdale, Lawrence Erlbaum Associates)

Jenkins, J.J. 1974. Remember that old theory of memory? Well, forget it! American Psychologist, November, pps. 785-795.

Lave, J. 1988. Cognition in Practice. Cambridge: Cambridge University Press.

Maes, P. 1990. Designing Autonomous Agents, Guest Editor. Robotics and Autonomous Systems 6(1,2) 1-196.

Schon, D.A. 1979. Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed), Metaphor and Thought. Cambridge: Cambridge Uni versity Press. 254-283.

Schon, D.A. 1987. Educating the Reflective Practitioner. San Francisco: Jossey-Bass Publishers.

Suchman, L.A. 1987. Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge: Cambridge Press.

Tyler, S. 1978. The Said and the Unsaid: Mind, Meaning, and Culture. New York: Academic Press.

Winograd, T. and Flores, F. 1986. Understanding Computers and Cognition: A New Foundation for Design. Norwood: Ablex.