Clancey,W.J. (1991) Why today's computers don't learn the way people do. In: P.A. Flach and R. A. Meersman (eds)*Future Directions in Artificial Intelligence*, Amsterdam: Elsevier. 53-62.

WHY TODAY'S COMPUTERS DON'T LEARN THE WAY PEOPLE DO

William J. Clancey
Institute for Research on Learning
2550 Hanover Street
Palo Alto, CA 94304

In *Future Directions in Artificial Intelligence,* P. A. Flach and R. A. Meersman, Ed. Amsterdam: Elsevier. 53-62, 1991.

A SYMPOSIUM, "WHAT CAN WE LEARN ABOUT LEARNING BY TEACHING MACHINES TO LEARN?" WAS ORGANIZED BY WALLACE FEURZIG AT THE ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION IN BOSTON (APRIL 1990). PRESENTATIONS WERE MADE BY OLIVER G. SELFRIDGE, ROGER C. SCHANK, AND MYSELF, WITH COMMENTARY BY LAWRENCE W. DAVIS AND ROBERT W. LAWLER. THE FOLLOWING IS AN ELABORATION OF MY PREPARED NOTES AND CAN BE TREATED AS AN EDITED TRANSCRIPT OF MY PRESENTATION.

The work being done at Schank's Institute for Learning Science emphasizing case-based inquiry, as well as the student programming projects at Papert's MIT Media Lab, is very exciting. But how are we to justify the constructionist approach (Papert, 1990)? If we follow the cognitive science view of the mind, which says that knowledge is stored as structures in memory, and learning is a matter of simply using and encoding the right structures, we might conclude that the right way to use computers for teaching is to build intelligent tutoring systems (Sleeman and Brown, 1982). Will these two alternative approaches--constructionism vs. instructionism, as Papert puts itcontinue to split the energy and resources of the research community, or can we explain why one is better than the other in view of how people learn, and go about our research in a more efficient way? One way to think of my work, which I will outline in the next fifteen minutes, is that it provides a new model of cognition--a new psychology--that supports a constructionist approach to teaching. In the short time available, I can only show you where I am headed; the ideas will perhaps sound crazy to some of you, but at least you'll know the research direction I advocate.

We could frame the question "Why computers don't learn the way people do" in different ways. We could consider the circumstances in which people learn, the materials available and social interaction. Instead, I've chosen to focus on the physical mechanism that supports what we as observers call learning. That is, I'm interested in how the brain works.

I'm going to begin by reading seven points, then I'll provide some background and elaboration on these ideas.

1. Computers don't learn the way people do because human memory is not a place where representations are stored.

2. Human learning doesn't consist of retrieving and applying structures, and then storing back modifications, which remain unchanged until their next use.

3. Knowledge does not consist of--cannot be reduced torepresentations, either descriptions of the world or descriptions of behavioral routines.

4. The stuff of AI programs--scripts, plans, strategies, grammars, schema hierarchies, semantic nets--are observer-relative descriptions of historical patterns, the product of interaction between an agent and an environment over time.

5. Such cognitive science representations--"knowledge-level theories"--are useful and *necessary*, but should not be identified with the mechanism inside human brains. They're necessary because we need to describe the combined system of people behaving in an environment, but that's different from describing the neurophysiological system inside individual heads.

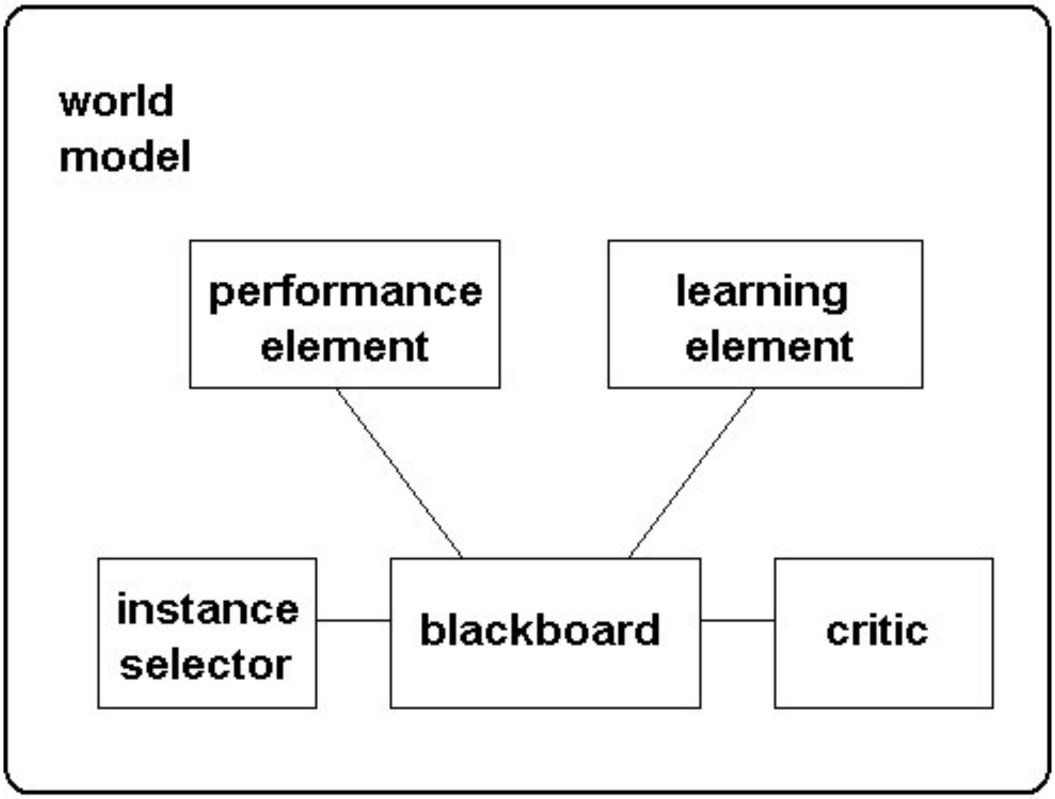
6. At heart, we've misunderstood the nature of representations. They are inherently perceptual--constructed by a perceptual process and given meaning by subsequent perception of them.

7. Computational theories of knowledge, memory, problem solving, and learning are fundamentally inadequate because the nature of perception is misconceived. Perception is not a peripheral process, but integrated as one process with behavior and learning.

Now, let's step back and reconsider how learning has been described in AI programs.

Here's how we described AI learning programs at Stanford's Knowledge Systems Laboratory in the late 1970s (Smith, et al., 1977).

THE COMPONENTS OF A LEARNING SYSTEM

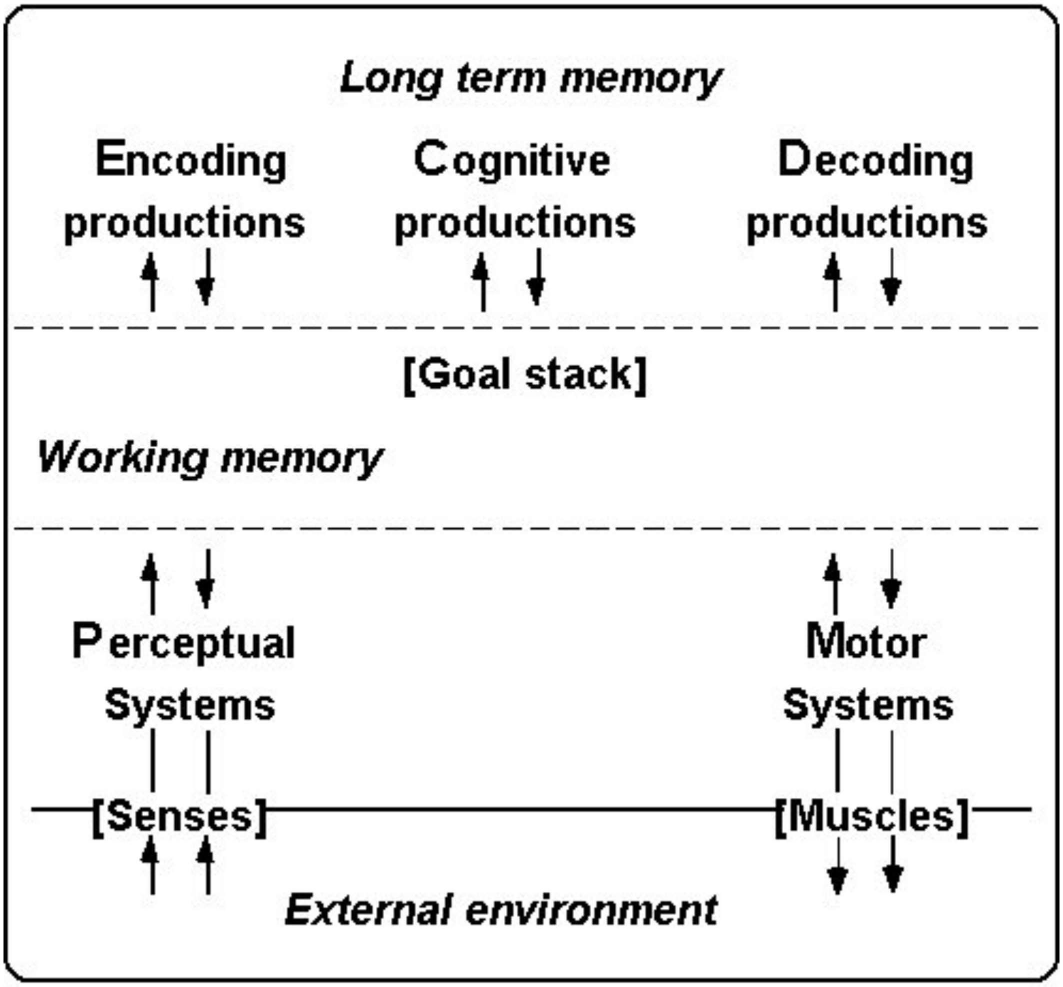


"A learning system responds acceptably with respect to some performance criterion within some time interval following a change in its environment."

Notice how the performance and learning elements are separate: According to the AI view, behaving and learning are distinct processes that take place at different times. Although some programs reorganize their knowledge during problem-solving to gain more efficient access or to extract what was only implicitly coded before, actual additions to the knowledge base are always made by a separate program. (To the authors' credit, they sought to make explicit assumptions about the environment, the "world model," as well as the constraints built into the critic.)

A more recent version of this diagram is Newell's generalization of SOAR in his book, "Unified Theories of Cognition" (in press).

TOTAL COGNITIVE SYSTEM



Performance:
[P ---> E] ---> C ---> [D ---> M]

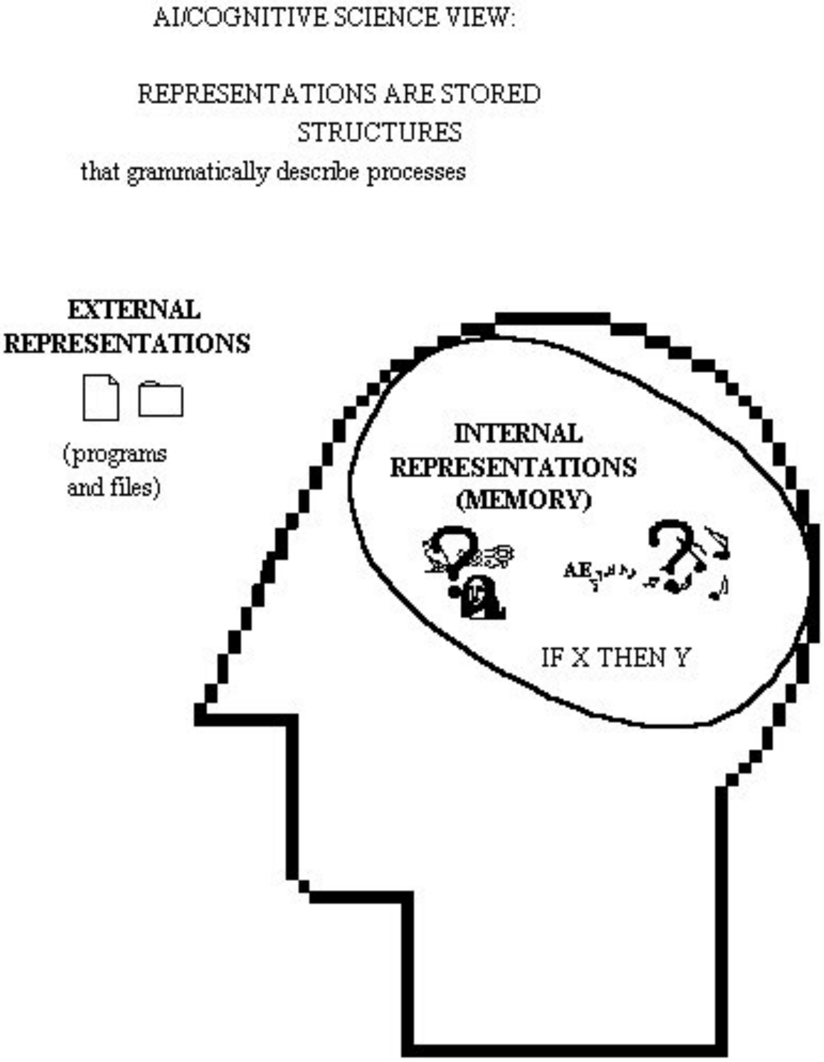
Structure and Learning:
[P] ---> [E ---> C ---> D] ---> [M]

This is intended to be a psychological model. But notice again how perceptual and motor systems are distinct from memory: Perceiving and moving are distinct from remembering, which goes on at a different time. Learning occurs in three places, as Newell shows on the bottom line.

According to this view, human cognition involves manipulating representations in a hidden way. The structures of working memory aren't always available for our conscious inspection. I claim that such structures, whether they exist or not, can't be representations. Representations must be perceived to be meaningful, that is, to be treated as representations.

Whether the structures are inside or provided as input, computer programs do not use representations at all in the sense that people do. Programs are only manipulating structures syntactically; they are not interpreting them, but only indexing their properties as in a database. The main error of AI and cognitive science has been to suppose that the interpretation of a representation is known prior to its production. But the meaning of a representation is neither predefinable nor static; it depends on the observer. Let me make this more explicit by another picture.

[At this point, I accidentally spill water on my slides...]



AI/COGNITIVE SCIENCE VIEW:

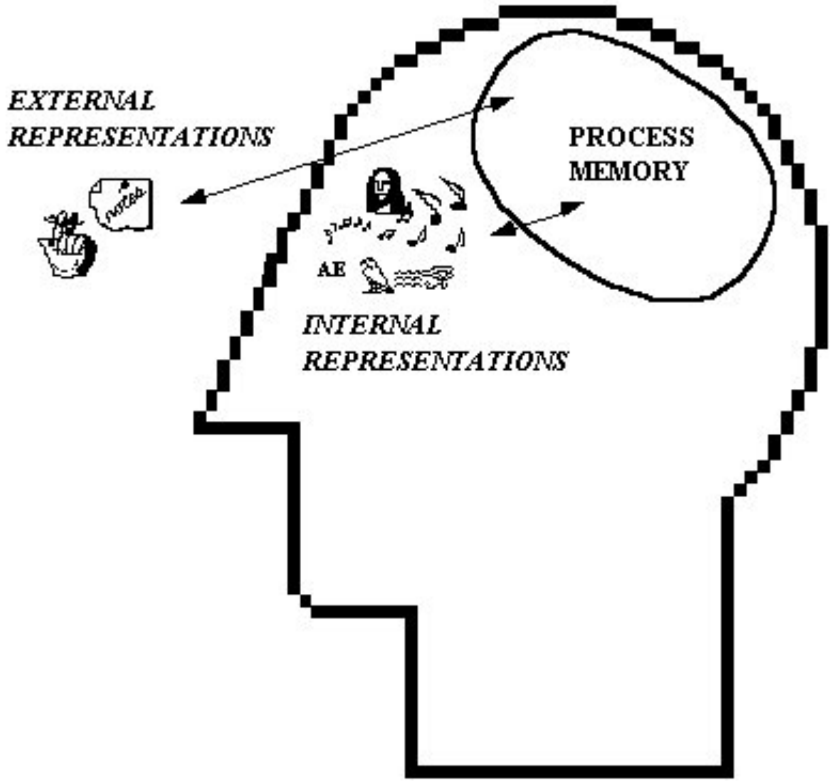
REPRESENTATIONS ARE STORED STRUCTURES that grammatically describe processes

This is the AI and cognitive science view of representations. As you can see, it is all wet. We have descriptions of processes written down in our programs and filesthis view claims that human memory is just the same, a place where descriptions are stored. So we expect to find strings like "IF X THEN Y" stored in human memory; it is indistinguishable from a knowledge base. Of course, we aren't sure what to do about pictures and sounds. That must be a different kind of memory, a different storage place.

Here's the alternative view I am developing.

REPRESENTATIONS ARE PERCEIVED:

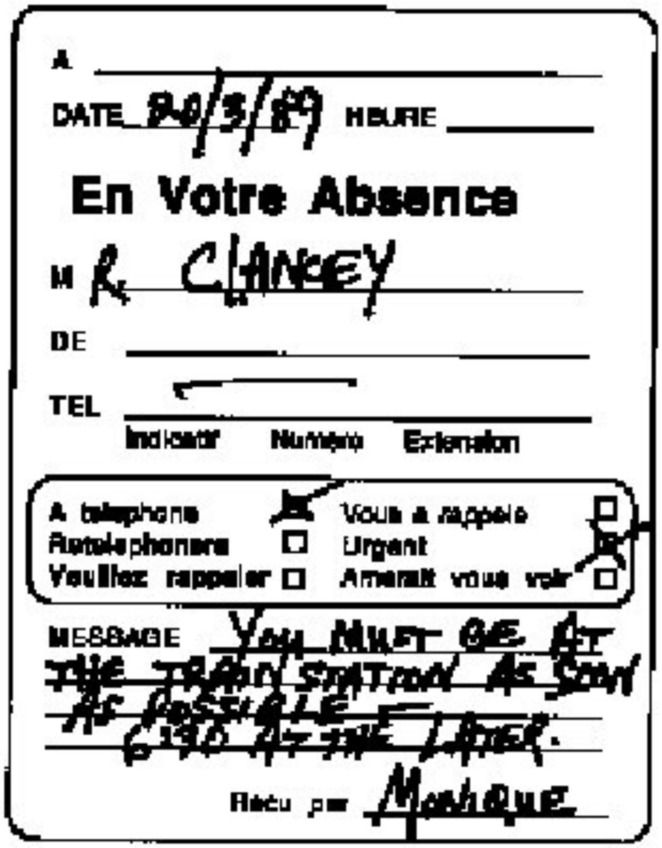
EXTERNALLY SENSED OR IMAGINED



I claim that memory is a capability to do things in ways similar to what we have done before, to reenact, sequence, and compose past behaviors. We have a memory for processes, not for descriptions of them. In Edelman's terms, these processes correspond to sensory-motor maps and maps of maps. Part of what we call memory is actually the conscious, cognitive process of constructing new sequences and composing them. Representations of course play a key role in orienting this process, but they must be perceived, either outside or imagined, such as silent speech or visualizations. Whatever is being manipulated internally, neurally, which we don't have access to, is not a representation. Getting clearer what representations are will help us better characterize the neural structures that *are* selected and reinforced subconsciously (following Edelman, I question the use of terms like "create" and "record").

I will illustrate these ideas by two examples. As Schank would say, let me tell you a story.

Here is a phone message I received at my hotel in Nice last March. We had just come in from dinner and maybe had a few drinks too many. "En Votre Absence: R. Clancey." Rosemary Clancey? Why did my mother call me? "You must be at the train station as soon as possible." What? (Paranoid thought: Somebody is forcing me to leave town!) "6:30 at the later." Tomorrow morning? Why?



A

DATE 20/3/89 HEURE

En Votre Absence

M R. CLANCEY

DE

TEL

Indicatif Numéro Extension

A téléphoné ☒ Vous a rappelé □
Retéléphonera □ Urgent ☒
Veuillez rappeler □ Aimerait vous voir □

MESSAGE YOU MUST BE AT THE TRAIN STATION AS SOON AS POSSIBLE 6:30 AT THE LATEST.

Reçu par Monique

It turns out that this message was supposed to be read to me over the phone, before dinner, while I was still in Antibes. It's a great example of the indexical nature of representations: How we interpret a representation, the way we talk about it, depends on our circumstances, including in this case the time of day, the city I was in, and whether I had been drinking. If I had heard this message before dinner, I would have known to go the Antibes train station and wait for my ride to take me to the restaurant.

This example may seem like an extreme case, but it points out the contextual nature of all representations: The meaning of a representation is not inherent in its form, but is constructed when we comment on the form. It's not a matter of *indexing* the meaning from memory (as Schank would have it), but *composing* it on the spot each time you reperceive what you take to be a representation. Again, a major error in cognitive science has been to suppose that the meaning of a representation is known prior to its production. It's the person perceiving the representation who determines what it means.

Here's another example. This picture was produced by Harold Cohen's program called AARON (Cohen, 1988).

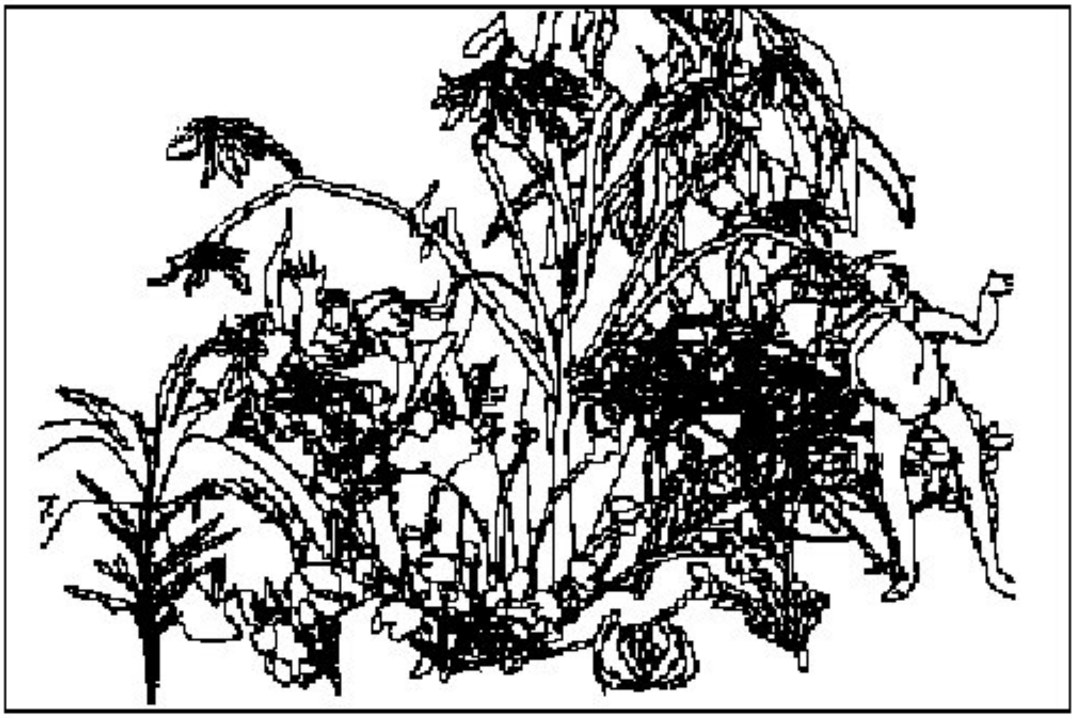


Figure 1: AARON drawing, 1987

Harold Cohen had a problem: He wanted AARON to produce original drawings, but how could this happen if Harold described the pictures ahead of time, when he wrote the program? If he used a grammar for describing the drawings, they would be Harold Cohen's drawings, not AARON's. In particular, how could he produce a drawing that looked three dimensional without putting three-dimensional descriptions of the world in the program ahead of time? The problem gets more complex if you imagine that AARON must produce a three-dimensional visualization in working memory before it draws. We're caught in a recursive conundrum: What produces the three-dimensional visualization if not another three-dimensional description?

What Harold discovered is that AARON could produce what an observer would take to be three-dimensional by viewing its drawings two-dimensionally, in terms of the placement of objects relative to the bottom of the drawing. In this way, the product (what observers perceive) and the mechanism are distinct. AARON draws by sensing and responding to local, two-dimensional features in its evolving drawing. In point of fact, AARON isn't literally looking at its drawing in the world, and Harold still built in grammatical descriptions of what trees and people look like (albeit as two-dimensional stick figures), but the overall approach is new. Its brings out the essential distinctions between production-mechanism and observer-perception, and suggests that mechanisms can be simpler than the descriptions observers make of the resulting behavior. Indeed, observer descriptions will have a global, historical quality that incorporates how the mechanism has interacted with its environment over time--which the mechanism doesn't necessarily need to produce its moment-by-moment behaviors. This is the idea behind the situated automata work of Brooks, Rosenschein, Agre, Steels, and others (Clancey, 1989a). Of course, an agent can reflect on his own behavior by objectifying it in a representation. These comments, in the form of goals, plans, and strategies, organize subsequent perception and hence organize the composition of new behaviors (Schon, 1987).

Again, the lesson is that we shouldn't try to build the interpretation of a representation into its production-mechanism. Rather, semantics are in the subsequent commentary, in the ongoing sequence of behavior, not in the individual structures or behaviors themselves. You can view what I have said about human memory and representations as a way of relating situated automata research to psychology.

Here's a summary of what I have said.

Why Don't Today's Computers Learn
the Way People Do?

o Memory is not a storage place for descriptions of how the behavior appears to an observer over time *(schemas, plans, grammars)*, but a capacity to directly reenact and compose previous ways of behaving.

o Learning is not a separate process, but an integral part of every perception and movement.

o Representations are not just syntactic patterns, but reperceived and given new meaning in every use by commentary that supplies a context for interpreting them.

+ Speaking is conceiving.

Memory is the name we give to the capability to behave in similar ways in similar situations. Perception does not require previous learning, as Rosenfield tells us in his book, "The Invention of Memory." Learning is not knowing what to do ("knowing how") plus knowing a theoretical description of the process ("knowing that"). That is, knowing what to do is not knowing two things (the lesson from Gilbert Ryle). Representations are not at the core of learning because memory is not a storage place for representations.

You can read more about these ideas in the *Proceedings of the Cognitive Science Meeting* (Clancey, 1989a) and *Machine Learning* (Clancey, 1989b). I hope you'll also read the books that have influenced me a great deal, including Bartlett's "Remembering," Gregory's "Inventing Reality: Physics as Language," Bateson's "Mind and Nature, a necessary unity."

As I said at the onset, I believe this view of memory and learning will support the constructionist approach. For example, you'll find that chapter 3 in Schon's book, "Educating the Reflective Practitioner" analyzes the evolving drawing of an architecture student in a way that's consistent with my view of how representations are given meaning and modified by commentary about them. Jeanne Bamberger's presentation (Bamberger, 1990), relating perception and talk about music, is similar.

Finally, I've put a statement at the bottom of the summary slide that I have found useful for keeping the central issues in focus. When we speak, we are not translating a description of what we are about to say. We are not manipulating grammars and producing internal strings that will be *decoded* (as Newell would have it) into motor commands. Speaking is conceiving. We don't know what we are going to say or what it might mean, until after we have created the representation. You might think about this when you are trying to recall the basic claims of my talk.

References

Bamberger, J. 1990. Logo music writer: Laboratory for intelligent music development. Annual meeting of the American Educational Research Association. Boston, MA.

Bartlett, F. C. 1977 (Original 1932). *Remembering-A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

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

Brooks, R.A. (In press). How to build complete creatures rather than isolated cognitive simulators. In K. vanLehn (editor), *Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition*. Hillsdale: Lawrence Erlbaum Associates.

Clancey, W.J. 1989. The frame of reference problem in cognitive modeling. *Proceedings of the Cognitive Science Society*, pps. 107-114.

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

Clancey, W.J. (in preparation). Review of Rosenfield's *Invention of Memory*. Submitted to the Journal of Artificial Intelligence.

Clancey, W.J. (in press). The frame of reference problem in the design of intelligent machines. In K. vanLehn (editor), *Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition* (Hillsdale: Lawrence Erlbaum Associates).

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.

Edelman, G.M. 1987. *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Gregory, B. 1988. *Inventing Reality: Physics as Language*. New York: John Wiley & Sons, Inc.

Papert, S., 1990, Introduction to Constructionist Learning. In I. Harel (editor), *A 5th Anniversary Collection of Papers Reflecting Research Reports, Projects in Progress, and Essays by the Epistemology & Learning Group,* The Media Laboratory, MIT, Cambridge, Massachusetts.

Rosenfield, I. 1988. *The Invention of Memory: A new view of the brain*. New York: Basic Books, Inc.

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

Sleeman, D. and Brown, J.S. 1982. *Intelligent Tutoring Systems*. London: Academic Press.

Smith, R. Mitchell, T., Chestek, R., and Buchanan, B.G. 1977. A model for learning systems. *Proceedings of the International Joint Conference on Artificial Intelligence*. Boston, MA, pps. 338-343.

Steels, L. 1990. Growing trees without a grammar. *Proceedings of the Second Artificial Life Conference*, Sante Fe, NM.