

Situated Cognition: Stepping out of Representational Flatland

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

A Response to Swann's Commentary, *AI Communications* (Vol. 4, No. 2/3, pp. 109-112, 1991).

I appreciate Philip Swann's effort in responding to the excerpts from my Delta-conference presentation (AICOM, Vol. 4 No. 1, March 1991, pps. 4-10). Unfortunately, Swann makes claims about situated cognition that have no basis in my talk or papers: "thought is only possible through language...there is no science between physiology and sociology...it [psychological models] never was a theory of anything...the entire biological and sensorial infra-structure of cognition is treated as irrelevant." Part of the problem is that the published transcript omitted my discussion of memory and perception (which is elaborated in the cited papers).

A second reason for miscommunication is that I am not merely presenting a point of view argued by other people—Dewey, Bartlett, and Wittgenstein, to name a few. Swann views my "Theory Y" as something "discovered in the library," which he characterizes as just an "unresolved conflict." He proceeds to burden me with all the misconceptions and confusions of other people. It is unfortunate that he chose to ignore the implications for education, which after all was the topic of the talk, and in itself demonstrates the novelty and value of the changed perspective. Although I credit other researchers, I am putting together something new, which I believe the following summary should make clear.

From the perspective of an AI researcher or cognitive scientist, situated cognition research can perhaps best be understood as the study of how representations are created and given meaning. An essential idea is that this process is perceptual and inherently dialectic. That is, the organization of mental processes producing coherent sequences of activity and the organization of representational forms (e.g., statements in a conversation, added lines to a sketch) arise together. A painter doesn't have a completed picture inside his head that he is merely executing on paper. A speaker in a conversation is not merely instantiating discourse plans and patterns. Even letters or words are not stored inside as descriptions of how they appear or how the hand or mouth moves. Mental organizations do not merely drive activity like stored programs, but are created in the course of the activity, always as new, living structures (Bartlett, 1932).

Understanding this requires a kind of Copernican shift: Representations are not at the center of the mind, but rather emerge from the interaction of mental processes with the environment. As Dewey (1902) pointed out, these "environmental" processes can be inside the head, as in silent speech and mental imagery. Representations are created so they can be perceived, externally in the form of speech, drawings, writing, or

directly experienced internally in the form of silent speech or visualizations. Representations are not structures stored, retrieved, and manipulated in a hidden way. Structures in the brain that cannot be perceived cannot be interpreted as being meaningful, so they have no representational status to the agent. We must distinguish between unperceivable neural structures, representational structures in the environment (e.g., texts, drawings, production rules), and representational experiences (e.g., hearing a word, visualizing something). This is the argument in Maturana's brilliant essay, *What is it to see?* (Maturana, 1983).

Therefore, in saying that cognition is situated, we mean that reasoning processes are not merely conditional on the environment, but are inherently brought into being during an *interactive process* (more precisely, an interaction of different systems, neural and environmental). Put somewhat metaphorically, the programs of the mind are not processing data, but are created during the process of representing what data needs to be processed. More precisely, data or "information in the environment" isn't merely described, selected, or filtered, but *constructed* in the course of perception (Bateson, 1972). Categorizations of things in the world are not merely retrieved descriptions, but created new each time (cf. Vygotsky (1934), "Every thought is a generalization.").

The adapted forms of intelligent behavior produced by the interaction between neural and environmental processes (e.g., strategies, habits, word senses, discourse conventions)—described by an observer over time from some frame of reference—cannot be ascribed to something pre-existing internally in neural structures (e.g., scripts, schemas, rules) or pre-existing in the world (e.g., properties of objects or events). The apparent static, object-like quality of behavior patterns follows from our interpretation of collected representations of behaviors (e.g., a schema-memory model, rules in a knowledge base, definitions in a dictionary). Seeing the static nature of representations of behavior, we come to think of descriptions as objective, talk about models as being "complete," and use phrases like "all the information in a situation"—as if interacting physical, biological, and social processes can be replaced by descriptions of their combined historical product.

Indeed, situated cognition leads us to reject both the idea that human memory consists of stored representations (i.e., descriptions of how behavior or the world appear to an observer over time) and the idea that reality has objective properties (Tyler, 1978; Lakoff, 1987). There is no *correspondence* between mental processes and the world because both our habits and what we claim to be true arise dialectically, by the interaction of mental processes and the environment. Concepts are not pre-defined feature lists stored like things in my head. I regenerate and reconstruct such representations in my acts of speaking, writing, drawing. I am not representing everything inside first and then translating descriptions of what I plan to say or do into overt behavior. When I do plan what to say or apply a grammar rule, I do it consciously, in *cycles* of perceiving a situation, articulating a rule, and reflecting on the rule.

To recapitulate, in my formulation, the situated cognition hypothesis is that *all processes of behaving, including speech, problem-solving, and physical skills, are generated on the spot, not by mechanical application of scripts or rules previously stored in the brain*. Representations are created and interpreted interactively, in cycles of perceiving and acting—representational forms are the *product of interactions*, the result of perceiving and behaving, not a fixed substrate from which perception or behavior is generated. The neural processes that lead to our speaking, drawing, gesturing, or any movement, are organized at the time of our behaving—we are not following scripts, applying rules, or executing procedures that describe how our behavior or the world appears. Following James, Bartlett, Gibson, Edelman, Rosenfield, and several other psychologists, I claim that human memory has no capability to store symbolic structures as structures (Clancey, in press). What we call memory is a capability to recompose and recoordinate ways of perceiving and acting. Higher mental processes (e.g., conversations, medical diagnosis, physics problem solving) are patterned because neural processes are biased to reorganize themselves, as perceptual categorizations and coordinated sequences of perceiving and acting.

Representations do not merely cause individual behaviors, which contribute on another level to social

interactions. The environment has its own emergent structures and patterns of interaction. Bartlett (1932) uses the example of a game like Rugby football:

"Nine-tenths of a swift game is as far as possible from the exploitation of a definite, thought-out plan, hatched beforehand, and carried out exactly as was intended. The members of the team go rapidly into positions which they did not foresee, plan, or even immediately envisage, any more than the bits of a glass in a kaleidoscope think out their relative positions in the patterns which they combine to make" (p. 277).

Bartlett goes on to say that if individuals have to think what another player is going to do, the team will be disconnected. This is not to say that we don't sometimes generate causal theories and plans, which subsequently change our behavior, but to underscore that behavior is often possible, indeed required, without them. This in turn is possible because, at the core, we always act directly, without referring to representations of the world or what we plan to do—perceiving and acting are dialectically coupled, not serial. This is not merely to acknowledge, as Swann suggests, that some mental behaviors are social in that they directly involve other people, but to make the fundamental claim that neural organizations come into being by interactions between neural and environmental processes. The emphasis on "social cognition" misses the point. Neural organizations arise in the course of activity, they are always new, they are not retrieved from storage. That is what the term "situated" means.

The memory-as-stored-structures hypothesis is tied to the idea in cognitive modeling and knowledge engineering that a representation of knowledge is knowledge itself. If I hold up a map of Europe, we all understand that it isn't Europe itself. But if I hold up a listing of rules from Mycin, we say, "Yes, that is knowledge itself." But knowledge cannot be inventoried. Knowing something is not having a thing, some substance in hand. The same is true of representations of meaning or context. Having in hand a representation of what a word means or of a situation is not understanding or being in a situation. Comprehending is not storing away a representation. Knowledge, like energy, is not a substance.

The idea that a knowledge base could be functionally equivalent to human capability fundamentally misconstrues the relation between *processes* and *pattern descriptions*. Collingwood (1938) says that we have misconstrued an experience of knowing, of comprehending, of discoursing with the product of the activity (the statements made). Having found patterns in these products (e.g., grammars, scripts, strategy rules), we supposed that these patterns were inside the head, constituting the mechanism that generated the observed behavior (Clancey, 1991b). It is true that most AI programs operate by storing and instantiating pattern descriptions, but that is not what people do.

For some areas of AI research, such as cognitive modeling and knowledge acquisition, the criticism of memory-as-stored-structures is of foundational importance. More generally, many AI researchers believe that a mechanism based on stored pattern descriptions can generate the range and flexibility of intelligent human behavior. Even if you don't agree with the claim that stored pattern descriptions can't explain how novel representations are created (recall Harold Cohen's dilemma in designing AARON), the claim about human memory is important because it suggests that other mechanisms are possible for producing intelligent behavior, which AI research has yet to exploit. In calling the memory assumption into question, I am trying to influence the course of AI research, and more generally how we conceive of people and how we design computer tools (cf. Winograd and Flores, 1986).

Unlike social scientists (e.g., Lave, Suchman) and other "nouvelle" AI researchers (e.g., Rosenschein, Brooks), I argue that we need a theory that relates to the neural structures of the individual person (Clancey, 1991b). By "neural" I mean especially perceptual processes that can be attributed to large organizations of neurons (e.g., Freeman's (1991) neural cell assemblies). By "social" I mean especially interactions between

people over the course of a few minutes. The study of how interacting neural and social processes—with their own multiple levels of coherent, emergent patterns of interaction—focus individual attention, bias perception, and organize movements is close to what Bartlett (1932) called *social psychology*.

Combining neural and social perspectives also involves explaining what is right about stored-schema models. I have taken pains to make this clear in a series of papers about the nature of knowledge-level models (Clancey, 1986; 1989a; 1989b; in press b). I stated explicitly in my *Frame of Reference* paper (Clancey, 1991b) that situated cognition must not throw out the baby with the bath water:

I am particularly concerned that we not lose sight of this modeling methodology as a legitimate, separate discipline, and to this end have recounted the main ideas in a series of papers. (p. 4)

In my Delta-conference talk I said,

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 modeling methodology. (p. 4)

Pattern descriptions now serve as a specification for how adapted behavior must appear, rather than the mechanism to be put inside the robot (Clancey, in preparation). Descriptions of novice-expert differences, reasoning strategies, explanation-based learning, etc. are descriptions of how people create and use models *within a representational language*, when interacting with their environment in cycles of perceiving and acting. To complement these descriptions, we need to understand *how representational languages are created*. Current psychological studies suggest this is a perceptual process—often involving collaborative, interpersonal interactions with physical materials—creating new forms, not interpreting a stored lexicon and grammar (Bamberger and Schön, 1983; Clancey and Roschelle, in preparation). Indeed, in people the model of the situation and the language for expressing it arise together in the course of interaction.

My long-term goal is to provide a mechanistic account of memory and learning at the level of brain processes that does justice to the interactional perspective (cf. Maturana, Gibson, Bateson). This account would provide an alternative basis for cognitive modeling, a reinterpretation of existing models, and a new mechanism for building intelligent machines. This is a tall order to flesh out. To make progress, cognitive scientists, AI researchers, and educators cannot continue to live in a representational flatland. Neither social nor neural science should be left to other researchers, as if they are merely levels of application and implementation for psychology (cf. Bransford, et al., 1977). The time is right for relating these perspectives, for creating a kind of neural-sociology of knowledge that will constitute a new cognitive science, which is neither individual nor social, but does justice to both. In the words of Dewey (1902),

It is easier to see the conditions in their separateness, to insist upon one at the expense of the other, to make antagonists of them, than to discover a reality to which each belongs....When this happens a really serious problem, that of interaction, is transformed into an unreal, and hence insoluble, theoretic problem.

References

- Bamberger, J. and Schön, D.A. 1983. Learning as reflective conversation with materials: Notes from work in progress. *Art Education*, March.
- 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.
- 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. 1986. Qualitative student models. In J. F. Traub (editor), *Annual Review of Computer Science* (pp. 381-450). Palo Alto: Annual Review Inc.
- Clancey, W. J. 1989a. The Knowledge Level Reinterpreted: Modeling How Systems Interact. *Machine Learning* 4 (1989) 287-293.
- Clancey, W. J. 1989b. Viewing knowledge bases as qualitative models. *IEEE Expert*, (Summer 1989), 9-23.
- Clancey, W.J. 1991a. Why today's computers don't learn the way people do. In P. Fläsch and R. Meersman (editors), *Future Directions in Artificial Intelligence*. Amsterdam: Elsevier, pps. 53-62.
- Clancey, W. J. 1991b. The frame of reference problem in the design of intelligent machines. In K. vanLehn (editor), *Architectures for Intelligence*, Hillsdale: Lawrence Erlbaum Associates.
- Clancey, W.J. (in press a). Review of Rosenfield's *The Invention of Memory*. To appear in the *Journal of Artificial Intelligence*.
- Clancey, W.J. (in press b). Model construction operators. To appear in the *Journal of Artificial Intelligence*.
- Clancey, W.J. (in preparation). A Boy Scout, Toto, and a bird: How situated cognition is different from situated robotics. A position paper prepared for the NATO Workshop on *Emergence, Situatedness, Subsumption, and Symbol Grounding*.
- Clancey, W.J. and Roschelle, J. (in preparation). Situated cognition: How representations are created and given meaning. To appear in *Educational Psychologist*.
- Collingwood, R. G. 1938. *The Principles of Art*, London: Oxford University Press.
- Dewey, J. 1902. *The Child and the Curriculum*, Chicago: University of Chicago Press.
- Freeman, W. J. 1991. The Physiology of Perception. *Scientific American*, (February), 78-85.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Maturana, H. R. 1983. What is it to see? *¿Qué es ver?* 16:255-269. Printed in Chile.
- Tyler, S. 1978. *The Said and the Unsaid: Mind, Meaning, and Culture*. New York: Academic Press.
- Vygotsky, L. (1934) 1986. *Thought and Language*. Cambridge: The MIT Press. Edited by A. Kozulin.

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