

A Boy Scout, Toto, and a Bird: How situated cognition is different from situated robotics

WJ Clancey
Institute for Research on Learning

In The "Artificial Life" Route to "Artificial Intelligence": Building Situated Embodied Agents, L. Steels and R. Brooks, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates. 227-236, 1995.

*A position paper prepared for the 1991 NATO Workshop
on Emergence, Situatedness, Subsumption, and Symbol Grounding*

We are at an exciting turning point in the development of intelligent machines. Situated robot designers (Maes, 1990) have given the AI community concrete examples of alternative architectures for coordinating sensation and action. These examples suggest that, for some navigation behaviors at least, predefined maps of the world and control structures are unnecessary. This work has developed in parallel with and lends credence to similar criticisms of models of human reasoning (Winograd and Flores, 1986; Suchman, 1987). However, it is crucial to understand that situated robotic designs are pragmatic, emphasizing engineering convenience and new ways of building machines. Brooks, et al. (1991) are not trying to model human beings, and to a significant degree their robotic designs violate situated cognition hypotheses about the nature of human knowledge and representation construction. I will sketch out some of these distinctions here, and suggest how they might be used to discover alternative architectures for robotics.

I believe that the fundamental question for robotic designers is how to construct an intelligent machine without bounding its behavior by the designer's preconceptions about the world (Clancey, 1991). By not building in maps and procedures that rigidly control behavior, situated robot designers seek more flexible, robust mechanisms, such that what the robot does develops in the course of historical interactions with the world. I have also argued that this research leads us to reconsider the relation of knowledge-level descriptions of behavior (an observer's descriptions of patterns in what the robot does over time in some environment) to the

mechanisms that coordinate sensation and action (e.g., a subsumption architecture designed by an engineer). I claim that a mechanism that reconstructs and re-coordinates *processes*, rather than stores and retrieves labeled descriptions or procedures, is more consistent with what we know about human memory and perception (Clancey, in press a; Clancey and Roschelle, in preparation). Such a process memory possibly cannot be built today, because we don't know how to build the kind of self-organizing mechanism that is required (cf. Freeman, 1991). But articulating how human cognition is different from a classical architecture helps delineate what aspects of situated robotic designs are still cast in the classical mold and remain to be freed of prevailing assumptions about the nature of memory and representations.

Situated Cognition Hypotheses

To begin, here are some of the hypotheses about cognition that essentially distinguish situated (human) cognition research from what Brooks et al. call “classical AI”:

- Knowledge is an explanatory concept like energy, a capacity for interactive behavior. Knowledge can be represented, but “knowledge is never in hand” (Newell, 1984). “The map is not the territory” (Korzybski, 1941).

Knowledge-level descriptions (e.g., prototype hierarchies, scripts, strategies) constitute an observer's model, characterizing patterns in behavior—the product of internal mechanisms and some environment—not structures or mechanisms inside the agent. Just as there is no such thing as “all in the information” in some situation, there is no such thing as completely describing an agent's behavior. Descriptions are always with respect to the frame of reference of an observer interacting with an agent-in-its-environment; descriptions embody the observer's point of view (including values and goals) and are themselves the product of interactions.

Cognitive models (including expert systems) replicate the patterns of human behavior—how it appears in recurrent interactions—without replicating the mechanism that produces human behavior. Such descriptions are necessary and valuable; they help specify what a cognitive architecture must be capable of accomplishing. In human behavior, such models, in the form of natural language grammars, disease hierarchies, operating procedures, etc. are extremely valuable for coordinating group behavior, or in general designing, controlling, diagnosing, and repairing complex systems (Clancey, in press b).

- Meaningful structures are not fixed, given, or static in either the environment or in human memory:

Human memory is not a place where things (e.g., schemas, categories, rules, procedures, scripts) are stored. Such representations—when they are not stored in the environment—are always constructed each time they are used. Representations are not manipulated by people in a hidden way, but must be perceived to be interpreted; that is, they must be in the environment (including silent speech and imagery). Interpretation is a process of commentary, constructing secondary representations that give meaning to experiences and perceptions by placing them in a context, thus relating them to activity (Suchman, 1987; Agre, 1988; Clancey, 1991).

Information is not given to the agent. Information is constructed by people in a process of perception; it is not selected, noticed, detected, chosen, or filtered from a set of given, static, pre-existing things (Maturana, 1983; Reeke and Edelman, 1988). Each perception is a generalization, a new construction. No category is merely retrieved or reinstated. In people every utterance is a new representation.

- **Human learning** occurs all the time. Every perception and coordinated movement is a generalization (Vygotsky, 1934), in the sense that it recomposes previous categorizations and sequences of behavior (Clancey, in press a). Perception and action are dialectic in people: What we perceive and what we say our perceptions mean arises together with what are doing and our sense of what we are doing (Schön, 1987; Bamberger, in press)

An important kind of learning occurs in cycles of behavior as we represent and comment on what we have done in the past (e.g., “explanation-based learning”). Knowledge-based approaches to machine learning model the learning that occurs in cycles of behavior, not the constant generalization that occurs with every action in people.

To summarize, human behavior is situated because 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. Knowledge can be represented, but it cannot be exhaustively inventoried by statements of belief or scripts for behaving. Knowledge is a capacity to behave

adaptively within an environment; it cannot be reduced to (replaced by) representations of behavior or the environment.

Representations are created by an *interaction* of neural and external processes in what we call perception. As the product of interactions with the environment (sensory, gestural, and interpersonal), representations cannot *correspond* to an external, objective reality. Representations are themselves interpreted interactively, in cycles of perceiving and acting—they are always outside the main loop; they are the *product* of interactions, not the physical substrate from which behavior is generated. Today's computer programs create and interpret representations grammatically, by applying patterns and rules. People construct a new representation with every interpretation.

Toto's Maps

We don't know how to design a machine today that respects our current hypotheses about human cognition. Situated robotic designs are valiant attempts to break away from past ways of programming, but they, perhaps necessarily, still embody many of classical AI's assumptions. For example, consider Mataric's robot dog, Toto (Mataric and Brooks, 1990; Mataric, 1991). Toto has an innovative design that enables it to learn the relative location of landmarks in some environment. But I would like to distinguish Toto's advances as a novel engineering design from its relation to situated cognition theory. To be brief, here is how Toto's design violates the hypotheses stated above:

Memory

Toto's design is based on predefined categories for modeling the world, such as wall, bearing, obstacle. *Descriptions* of landmarks (e.g., "leftwall," compass bearing) are stored in a graph during Toto's operation.

Learning

Toto uses the classical approach of comparing the current landmark to a stored description of type, bearing, and position. This matching process uses a predefined calculus for manipulating the representation, just as in rule-based systems. For example, the calculus represents the equivalence of a left wall heading south and a right wall heading north (Mataric and Brooks, 1990). Toto doesn't learn with every interaction; for example, it doesn't update its graph if an obstacle isn't a known landmark.

In summary, Toto adheres to classical views about information as given, memory as description storage, and learning as controlled, grammatical manipulation of descriptions.

On the other hand, Toto's design is consistent with and indeed motivated by the view that knowledge-level descriptions of behavior (e.g., wall following) needn't be encoded in the mechanism as a map of the environment and fixed procedure for moving about. The fact that Toto constructs a map is not novel in itself. What is new and especially interesting is how Toto stores the map and how map-building is coordinated with primitive behaviors. In particular, the map is not globally available. Stored information is only accessible in the context of moving through the environment, when the history of interactions activates the nodes in the landmark graph. Furthermore, the graph is dynamically created as "jumper links," so that landmark recognition activates the next landmark detection process. This effectively replicates the "next-next-next" nature of human memory, what Bamberger calls the "felt path" (Bamberger, in press).

The separation of the map from the motion and sensing behaviors also appears to be a good idea, in so far as we view the map as an internally constructed representation that other processes apprehend and respond to (in the manner of Minsky's B and A brains (Minsky, 1986)). My complaint, however, is that descriptions of current landmarks and a sense of similarity with past categorizations should be *co-constructed* with the robot's high-level coordination of its primitive behaviors (reflex movements). That is, how "what is out there" is categorized should arise with the process of categorizing "what the robot is doing now." As it stands, the design violates Brook's own principle that perception is not an input to action—Mataric and Brooks have simply moved the serial, left-to-right precedence to a serial, bottom-to-top precedence.

The claim of situated cognition (in my formulation) is that perception and action arise together, dialectically forming each other. Perceiving landmarks is not retrieving past descriptions and matching against current categorizations (Maturana, 1983; Schön, 1979). In the human, there is no structure stored in the past to compare the present to (Bartlett, 1932; Gibson, 1966; Reeke and Edelman, 1988). Toto's "active representations" graph *models* the process of activation by which processes of past perceiving and moving are coordinated, but descriptions of past encounters are *stored*. In people the processes themselves are literally reconstructed by reactivating neural nets that actually do the coordination (the sensing and the moving), not nets

that store descriptions. Simply put, the claim is that people navigate through familiar space without referring to representations; sensations are directly coupled to actions without intermediate *acts* of description. In comparing and disambiguating *descriptions*—“the landmark I am sensing now” and “the landmark description I stored in my graph”—Toto is simulating *reasoning*, which is more complex behavior than we expect to see in a model of a dog.

This brief analysis illustrates that we need conventions for describing alternative robotic mechanisms, so we can better describe what is new and what work remains to be done. What distinguishes situated robotics and classical AI is muddled because how classical programs work has been poorly articulated relative to our current needs. Useful concepts can be derived by first comparing classical programs to situated cognition hypotheses; this gives us comparative descriptions like “memory-as-structure storage vs. memory as a capacity for recomposing past coordinations” and “learning via perceptual generalization (within a cycle; what people must do because they don’t store representations) vs. learning via grammatically manipulating representations (in cycles of perception and action, e.g., explanation-based learning in machines).” The most glaring problem is that how people create and use representations has been almost universally misconstrued in classical AI (Clancey and Roschelle, in preparation). Situated robotics has yet to address how coordination of sensation and action in complex spaces or in *sequences of behavior* over time is reconstructed, without storing descriptions of either behavior or the world (Rosenfield, 1988).

We must distinguish representations used by people (road maps, journal papers) from assumed structures in the head that aren’t perceivable. The processes of constructing and interpreting representations that occurs in cycles of human behavior is radically different from hidden manipulation of neural structures. To call both perceived structures like maps and unperceivable neural structures “representations” is to confuse what intelligence is. In this respect, Toto models how people use coordinate systems in *cycles* of behaving.

Situated cognition theories suggest that representations don’t mediate human behavior within each cycle (Winograd and Flores, 1986); in particular, we can walk through a room *without referring to an internal map of where things are located*, by directly coordinating our behaviors through space and time in ways we have composed and sequenced them before (a process memory, cf. Rosenfield, 1988). It is bizarre to postulate that dogs represent what people get by quite well without, and

even more strange to assume that dogs have developed coordinate representational *languages* (e.g., “bearing,” “left-wall orientation”). Indeed, how a dog could want to go *somewhere* (a particular place or kind of place) without having a descriptive language? The situated cognition claim is that the coordination is accomplished in dogs by reactivation of past neural compositions (sensory-effector maps and maps of maps producing sequences of behavior).

We must distinguish more carefully between what it means for a Boy Scout to use a compass bearing, what it means for Toto to store descriptions of landmarks, and how birds might migrate by interacting with a magnetic field (Baker, 1981). I claim that the Boy Scout is more like the bird than Toto, because he doesn’t literally store descriptions. It may be tempting to say that a process memory enables the same behavior as structure storage (e.g., the Boy Scout can say, “I remember that its bearing was 45 degrees”). But this again confuses how behavior appears with the flexibility and generative capabilities of different architectures.

Recommendations

To proceed effectively and systematically, robot designers and their critics might concentrate on the following:

1. Be clear what design alternatives you are using and why. Speak in terms of memory, perception, learning. What representations of the world are built in? What is stored? How are sensation and action coordinated? How are routines learned?

Attempt to develop a language for classifying systems:

- a) categorical perception vs. only direct sensation.
- b) maps, map primitives, or grammars for creating maps are hardwired.
- c) composite behaviors (e.g., sentence templates), primitive behaviors (e.g., reflexes), or constraints between behaviors are hardwired.
- d) opposing behaviors built in (e.g., left and right turn); sensors are fixed or mobile.

2. Experimentation:

- a) state hypotheses and purpose of the experiment (not just its design).
- b) change the environment systematically, and justify your choice of a microworld.
- c) experimentally explore and describe surprises, but work within a framework that defines a space of experiments.

3. Specify a robot’s behavior using classical representations (e.g., scripts, grammars, situation-action rules) so we can compare the capacities or “knowledge” of different

designs (including after learning). Similarly, specify environmental assumptions using classical representations (e.g., quantitative and qualitative models). Principled robot design requires systematically describing behaviors and environments.

4. Define the enterprise in terms of specific constraints:

- a) Functional: Are you designing for a particular environment? For particular learned behaviors?
- b) Biological: Are you replicating animal capacities?
- c) Computational: Are you doing a bottom-up experiment to see what a given mechanism can do?

5. Don't view the design of a society of robots as a different research problem. Ignoring the effect of other agents is just a variation of ignoring how the environment can structure behavior (presumably the view we are arguing against). Look for ways that emergent multi-agent patterns of interaction can be perceived by individuals and structure individual behavior (Steels, 1990).

6. Move towards construction of processes, not just activation of prewired constraints between behaviors. Move from the idea of predetermined, layered control (subsumption architecture) to creating new compositions (literally new networks) that can be reactivated (and potentially generalized rather than simply re-enacted). Programs like Toto, compared to people, are both too reactive (no learning of procedures, composite behaviors that effectively become new primitives) and too predetermined (no learning of categories, new ways of coordinating behaviors outside the subsumption layering). Correlating multi-modal sensation might be a practical and not too complex starting point.

Conclusions

We shouldn't expect progress to be monotonic; we need to take a broad view of the difficulty of articulating what we are doing. For example, some might view Winograd's early work (SHRDLU) as a mistake, particularly in light of his subsequent rejection of that approach. But progress requires clearly and valiantly pushing a point of view, so the community can reflect on it and see where it falls short. In this respect, Mataric's design of Toto is a major contribution to AI, and especially valuable as a foil for explaining situated cognition hypotheses. With such artifacts in hand, we can say better what we have done, what we are trying to do, and what to try next. Given the tentativeness of our theories and the compromises inherent in our engineering designs, we would be well-advised to retain some humility—looking back a few years (or even months) from now, we may realize that

we've made the same mistakes as classical AI. Before we proclaim that the path through the desert is now found, we should remember that it is unlikely that any trend or school of thought—whether behaviorism, gestalt psychology, or classical AI—is entirely wrong.

Acknowledgement

I am grateful to Maja Mataric for providing useful explanations of Toto's design, as well as thoughtful suggestions for improving these comments.

References

- Agre, P. 1988. The dynamic structure of everyday life. Dissertation in Electrical Engineering and Computer Science, MIT.
- Baker, R. 1981. *The Mystery of Migration*. New York: Viking Press.
- Bamberger, J. in press. The Mind Behind the Musical Ear.
- Brooks, R. 1991. Intelligence without reason. *IJCAI Proceedings*. Sydney, Australia.
- Clancey, W. J. 1991. 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. and Roschelle, J. (in preparation). Situated Cognition: How representations are created and given meaning. Presented at AERA91, Chicago. To appear in a special issue of the *Educational Psychologist*.
- Freeman, W. J. 1991. The Physiology of Perception. *Scientific American*, (February), 78-85.
- Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin Company.
- Korzybski, A. 1941. *Science and Sanity*. New York: Science Press.
- Maes, P. 1990. Designing Autonomous Agents, Guest Editor. *Robotics and Autonomous Systems* 6(1,2) 1-196.
- Mataric, M. and Brooks, R. A. 1990. Learning a distributed map representation base don navigation behaviors. USA-Japan Symposium on Flexible Automation.

- Mataric, M. 1991. Behavioral synergy without explicit integration. Proceedings AAAI Spring Symposium on Integrated Intelligent Architectures, to appear in SIGART.
- Maturana, H. R. 1983. What is it to see? ¿*Qué es ver?* 16:255-269. Printed in Chile.
- Minsky, M. 1986. *The Society of Mind*. New York: Simon and Schuster.
- Newell, A. 1984. The knowledge level, *Artificial Intelligence* **18**(1) 87-127 .
- Reeke, G.N. and Edelman, G.M. 1988. Real brains and artificial intelligence. *Daedalus*, **117** (1) Winter, "Artificial Intelligence" issue
- Rosenfield, I. 1988. *The Invention of Memory: A New View of the Brain* New York: Basic Books.
- Schön, D.A. 1979. Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed), *Metaphor and Thought*. Cambridge: Cambridge University Press. 254-283.
- Steels, L. 1990. Cooperation through self-organization.
- Suchman, L.A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge 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.