AI: Inventing a New Kind of Machine

WILLIAM J. CLANCEY

Institute for Research on Learning, 2550 Hanover Street, Palo Alto, CA 94304
⟨bill_clancey@irl org⟩

During the 1980's, related hypotheses about cognition emerged in computer science and neuroscience:

(1) AI researchers began to realize that knowledge engineering—the process of codifying human knowledge in expert systems—makes a solid contribution to computer modeling, but such programs neither replicate nor explain the full capability of human intelligence.

(2) Exclusively "symbolic" theories of cognition inadequately distinguish between an observer's *description of patterns* in intelligent behavior and the underlying mechanism that relates sensory and motor processes in biological systems. Symbolic theories view *reasoning* in words and other symbols as both the product and engine of thought.

(3) Word-based representations such as rules and scripts enable "knowledge-based" programs to effectively model and control complex systems in the world. But a mechanism built only out of networks of words and their relations is limited in its ability to conceive and coordinate behavior in multiple sensory modalities in time. Expert systems cannot dance or smell a change in the weather.

(4) A means-ends approach to engineering an intelligent machine now suggests that we focus on the differences between biological processes and symbolic models—not to replicate human thinking per se, but to replicate how a neurological system dynamically relates sensation, perceptual categorization, conceptual coordination, and movement. Studies of cognition should be broadened to understand simpler mechanisms than the human brain, especially by considering path finding in insects, birds, and other mammals.

Initial studies of differences between symbolic models and neural processes suggest two basic limitations in the traditional approach of game-playing programs, expert systems, natural language processing, automated planning, and so on. First, human memory is much more than a storehouse where structures are put away, indexed, and retrieved by rote. Second, human reasoning involves more than searching, matching, and recombining previously stored descriptions of situations and action plans. Indeed, these hypotheses are related: remembering and reasoning both involve *reconceptualization*. That is, human learning occurs with every thought; learning is not just a reflective process of formalizing knowledge after deeds are done.

By this view, often called "situated cognition," the mechanisms of remembering, learning, perceiving, and moving are integrated in ways that today's computer hardware and software do not replicate [Clancey 1993]. The brain's relation of sensation, conception, and movement is more flexible in dynamic recombination, sequencing, and generalization than the best symbolic models. Explorations of parallel processing and connectionism begin to articulate these distinctions. However, specialists in neurobiology and neuropsychology argue that although intelligent behavior is obviously serial in time and neural modules may operate in parallel, the temporal and spatial dependence of neural processes is different in kind from the control structures of either the traditional von Neumann architecture or the parallel-independent form of distributed processing.

Hence, for many computer scientists human intelligence—the phenomenon to be replicated—and knowledge are placed in a new relation. Previously, we explained intelligent behavior in terms of knowledge stored in memory: word definitions, facts and rules modeling complex causal processes in the world, plans and procedures for doing things. Now we view these as *descriptive models*—which people use like maps, handbooks, and standards for guiding their actions. These descriptions, like the rules in an expert system, are not knowledge itself but *representations of knowledge*—a step removed from what is occurring in the brain. Like an article in an encyclopedia, such descriptions require interpretation in the context of use, which in people involves not only generating additional descriptions but conceptualizing in nonverbalmodes. As Newell put it, "Knowledge is never actually in hand. . . . Knowledge can only be created dynamically in time."

How are these new hypotheses pursued by AI researchers today? In applying knowledge engineering, for example, researchers are more aware that model building is not merely a process of extracting descriptions from the expert's memory, but of constructing new descriptions, new models of physical, biological, and social phenomena. In cognitive modeling, our research focus shifts to the interactional dynamics of sensory categorization and coordinated motion. For example, natural language processing in Soar now examines how working memory is not just a buffer for temporary storage, but a constructive process by which a new motor sequence is organized [Polk and Newell 1995]. In robotics, a subfield of "situated automata" revisits the cybernetic concern with feedback and emergent effects from the combination of different mechanisms [Steels and Brooks 1995]. A now classic example is Brooks' wall-follower, which combines a sideward-repelling action with a forward motion.

The change in perspective can be bewildering: no longer are symbols, representations, text, concepts, knowledge, and formal models equated. Symbols in the brain are not immutable patterns but structural couplings between modular systems, continuously adapted in use by a generalization process that dynamically recategorizes and sequences different modalities of sensation, conception, and motion. Representations are not just coded artifacts such as diagrams and text, but in the brain are active processes that categorize and sequence behavior in time [Edelman 1992; Rosenfield 1992; Bickhard and Terveen 1995]. Conceptualization is not just verbal but may be rhythmic, visual, manipulospatial—ways in which the neural system categorizes sequences of multimodal categorizations in time. In the view of Howard Gardner [1985], people have multiple intelligences, with different ways of making sense and coordinating action. The neuropsychological studies of Oliver Sacks [1987] reveal how people who cannot verbally abstract their experience may nevertheless dance and speak poetically, and people who speak in the particulars of geometric forms and rules may be unable to see the forest for the trees. Indeed, our robots, like these patients, once loaded with a knowledge base and shoved out into the world of shopping malls and national parks, may appear to be dysfunctional morons or *idiots savants*.

With these observations in mind, we seek to invent a new kind of machine—a designed artifact—with the modularity and temporal dependence of neural coordination. Some connectionists are exploring computational recognition systems based on dynamic coupling, in which subsystems exist only in mutual reinforcement [Freeman 1991]. Structures in the brain form during action itself, like a car whose engine's parts and causal linkages change under the load of a steeper hill or colder wind. Such mechanisms are "situated" because the repertoire of actions becomes organized by what is perceived. Again, the thrust is not to "understand people" or to "replicate the details of human thinking," as many researchers view a psychological approach. Rather, the focus on biological flexibility and development is more fundamental—to understand the mechanism of *dynamic recoordination* in neurological systems, which apparently both reuses previous system configurations and generalizes them in real time.

In retrospect, we view symbolic models (e.g., natural language grammars) as descriptions of gross patterns that "subsymbolic" systems produce in the agent's interactive behavior. The "symbolic" AI of the 1950s through the early 1980s is not wrong, but it may be a bit inside out. People really do articulate rules, scripts, and definitions and refer to them. These descriptive representations are not in their brains, but on their desks and in their libraries. The representations in the brain are different in kind—dynamically formed, changing in use, *categorizations of categorizations over time*, and multimodal (not just verbal), as means of *conceptual coordination*. Meaning can be described, but words are not concepts; the map is not the territory.

To some researchers, this entire shift is misconceived. For others, it is just part of an ongoing conflict of different views: the public debate in the 1930s between Bertrand Russell and John Dewey, the argument among neuropsychologists since the 19th century about how to explain specific memories elicited by neural probes, and the nature-versus-nurture controversy over Chomsky's universal grammar. Our discussion is complicated by suspicion about "armchair" philosophy, shifting interpretations about past research (of course, representations could be more than encodings, they say now, when in the 1970s there was "only one game in town"), and the natural diversity and separation of different schools and projects within even a narrow discipline (who has read the recent two *volumes* of Soar papers?).

Predicting AI results is a notoriously hazardous act, but even a staunch critic would have to agree that the uproar of the past decade is a sign not of a sleeping field but of a vibrant, even outrageously changing community. Full of new ideas, we are heading in several directions at once, rededicated to understanding the living biological processes that inspired us to create computers in the first place.

REFERENCES

BICKHARD, M. H. AND TERVEEN, L 1995 *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. Elsevier Science, New York.

CLANCEY, W. J. 1993. Situated action: A neuropsychological interpretation (Response to Vera and Simon). *Cognitive Sci. 17*, 1, 87–107.

EDELMAN, G. 1992 *Bright Air, Brilliant Fire. On the Matter of the Mind*. Basic Books, New York.

FREEMAN, W J. 1991. The physiology of perception. *Sci. Am.* (Feb.), 78–85.

GARDNER, H. 1985. *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, New York.

POLK, T. A. AND NEWELL, A 1995 Deduction as verbal reasoning. *Psychol. Rev.* (in press).

ROSENFIELD, I 1992 *The Strange, Familiar, and Forgotten* Vintage Books, New York.

SACKS, O 1987. *The Man Who Mistook His Wife for a Hat*. Harper & Row, New York

STEELS, L. AND BROOKS, R., Eds. 1995. *The "Artificial Life" Route to "Artificial Intelligence": Building Situated Embodied Agents*. Lawrence Erlbaum Associates, Hillsdale, NJ.