The Frame of Reference Problem in Cognitive Modeling

William J. Clancey

Institute for Research on Learning

ABSTRACT

Since at least the mid-70's there has been widespread agreement among cognitive science researchers that models of a problem-solving agent should incorporate its knowledge about the world and an inference procedure for interpreting this knowledge to construct plans and take actions. Research questions have focused on how knowledge is represented in computer programs and how such cognitive models can be verified in psychological experiments. We are now experiencing increasing confusion and misunderstanding as different critiques are leveled against this methodology and new jargon is introduced (e.g., "not rules," "ready-to-hand," "background," "situated," "subsymbolic"). Such divergent approaches put a premium on improving our understanding of past modeling methods, allowing us to more sharply contrast proposed alternatives. This paper compares and synthesizes new robotic research that is founded on the idea that knowledge does not consist of objective representations of the world. This research develops a new view of planning that distinguishes between a robot designer's ontological preconceptions, the dynamics of a robot's interaction with an environment, and an observer's descriptive theories of patterns in the robot's behavior. These frame-of-reference problems are illustrated here and unified by a new framework for describing cognitive models.

CHANGE AND CONFLUENCE IN COGNITIVE SCIENCE

What accounts for the regularities we observe in intelligent behavior? Many cognitive scientists would respond, "Mental structures which are representations, symbols of things in the world." Since at least the mid-70's there has been widespread agreement among cognitive science researchers that models of a problem-solving agent should incorporate knowledge about the world and some sort of inference procedure for interpreting this knowledge to construct plans and take actions. Research questions have focused on how knowledge is represented in computer programs and how such cognitive models can be verified in psychological experiments.

We are now experiencing a change in the dominant paradigm, as different critiques are leveled against this methodology and new computational models, based on the idea of neural networks, are introduced. There have been many philosophical arguments against Cognitive Science and AI research over the years; what reason is there to suppose that we are making progress now on these complex issues? Most striking is the convergence of ideas and new approaches over the past 5 years:

- The long-standing criticism of Dreyfus (1972), for example, has been joined by insiders (Winograd & Flores, 1986)(Clancey, 1987)(Rommetveit, 1987);

- Neural net research has reminded us of the extent of the gap between neurobiology and cognitive science models, while new hardware and programming techniques have enabled a resurgence of network modeling (Rumelhart, et al.,1986);

- Cognitive science itself has flourished and succeeded in including social scientists within the community, and their methods and analyses often starkly contrast with the AI view of human knowledge and reasoning (Suchman, 1987). For example, increasing emphasis is placed on representation construction as an activity within our perceptual space, organized by social interaction, (e.g., Allen, 1988), not something in our memory that precedes speaking, drawing, or action in general.

Criticisms of AI and cognitive science may have often failed to be effective because they aren't sufficiently grounded in computational modeling terminology and may even appear to be compatible with existing programs. For example, the current buzzword "situated" might just mean "conditional on the input data of particular situations"; hence all programs are situated. The discourse of another intellectual tradition may even appear incoherent to cognitive scientists; consider for example, "representation must be based on interactive differentiation and implicit definition" (Suchman, 1987, p. 78). Experienced AI researchers know that an engineering approach is essential for making progress on these issues. Perhaps the most important reason for recent progress and optimism about the future is the construction of alternative cognitive models as computer programs, the field's agreed basis for expressing theories:

- The AI-learning community is focusing on how a given ontology of internal structures--the designer's prior commitment to the objects, events, and processes in the world--enables or limits a given space of behavior (e.g., the knowledge level analyses of (Dietterich, 1986) (Alexander, et al., 1986)).

- New robots ("situated automata") demonstrate that a full map of the world isn't required for complex behavior; instead, maintaining a relation between an agent's perceptual state and new sensations enables simple mechanisms to bring about what observers would call search, tracking, avoidance, etc. (Braitenberg, 1984)(Brooks, 1988)(Agre, 1988) (Rosenschein, 1985).

- Neural networks, incorporating "hidden layers" and using back-propagation learning, provide a new means of encoding input/output training relationships, and are suggestive of how sensory and motor learning may occur in the brain (Rumelhart, et al., 1986).

In essence, this new research reconsiders how the internal representations in an agent derive from the *dynamics of a physical situation*, relegating an observer's later descriptions of the patterns in the agent's behavior (what has been called "the agent's knowledge") to a different level of analysis. That is to say, this new research suggests that we reclassify most existing cognitive models as being *descriptive and relative to an observer's frame of reference*, not structure-function mechanisms internal to the agent that cause the observed behavior.

By systematically analyzing these emerging alternative intelligent architectures, placing them in ordered relation to each other, we should be able to articulate distinctions that the researchers couldn't accomplish alone. The result will be a better understanding of the diverse approaches to "situated cognition" and "neural networks" research, contrasted against conventional AI research. Thus, understanding a new approach and reconceptualizing what a "traditional" approach was about will arise together.

My approach here is to characterize the ontological commitments of the alternative models: What facts about the world are built into each program? Two useful, related questions are: *Who owns*

the representations (robot, designer, or observer)? *Where's the knowledge* (in a designer's specification, in robotic memory, in the relation of the robot to its environment, or in our statements as observers)? Throughout, I will use the term "robot" to emphasize that we are dealing with designed artifacts intended to be agents in some known environment. I believe we need to distance ourselves from our programs, so we can better understand our relation to them. "Robot" here means any cognitive model implemented as a computer program, specifically including computational models of people. Our orientation here is not of philosophical discourse in the abstract, but rather trying to find *an appropriate language to describe existing robots and the process by which they are designed,* so the engineering methods for building them are clear enough to order, compare, and improve upon them. In the conclusion, I will reach beyond what has been currently built in order to articulate what designers are attempting to achieve and to relate to other analyses in anthropology, linguistics, and philosophy.

THE PROBLEM: THE ONTOLOGICAL COMMITMENTS OF PLANS

When we examine situated automata research, we find a striking emphasis on the *nature of planning*, focusing on the ontological commitments made by the designer of the computer program. Agre and Kaelbling (Kaelbling, 1988) emphasize the resource and information limitations of real-time behavior--deliberation between alternatives must be extremely limited and many details about the world (e.g., will the next closed door I approach open from the left or the right?) can't be anticipated. Rosenschein found that formal analyses of knowledge bases are problematic--how can they be related in a principled way to the world, when their meaning depends on the designer's changing interpretations of the data structures? Cohen was wedged in a designer's conundrum: Since AARON is supposed to be producing new drawings of people standing in a garden, how could he build in a representation of these drawings before they are made? (Cohen, 1988) Cohen was face to face with the ultimate ontological limit of traditional cognitive models: Any description of the world that he builds in as a designer will fix the space of AARON's drawings. How then can a robot be designed so it isn't limited by its designer's preconception of the world? If such limitations are inevitable for designed artifacts, how can the specification process be accomplished in a principled way? Following are four perspectives on these questions.

Classical Planning -- Knowledge is in the robot's memory

In most AI/cognitive science research to date, the descriptions of regularities in the world and regularities in the robot's behavior are called "knowledge" and located in the robot's memory. A robot preferably uses a declarative map of the world, planning constraints, metaplanning strategies, etc. This view is illustrated especially well by natural language programs, which incorporate in memory a model of the domain of discourse, script descriptions of activities, grammars, prose configuration plans, conversational patterns, etc. Aiming to cope with the computational limits of combinatoric and real-time constraints, some researchers are re-engineering their programs to use parallel processing, partial compilation, failure and alternative route anticipation, etc. These approaches might incorporate further ontological distinctions (e.g., preconceptions of what can go wrong), but adhere to the classical view of planning.

Knowledge is in the designer's specification.

Rosenschein introduces an interesting twist. Besides using efficient engineering (compiling programs into digital circuits), his methodology explicitly views the robot as a designed artifact. He formally specifies robotic behavior in terms of I/O and internal state changes, gaining the advantages of internal consistency and explicitly-articulated task assumptions. The problem of

building a robot is viewed as an engineering problem, nicely delineating the designer's relation to the robot and the designed behavior.

Knowledge is not incorporated as data structure encodings; it is replaced by a design description that specifies how the state of the machine and the state of the environment should relate. Thus, knowledge is not something placed in the robot, but is a theoretical construct used by the designer for deriving a circuit whose interactive coupling with its environment has certain desirable properties. These "background constraints . . . comprise a permanent description of how the automaton is coupled to its environment and are themselves invariant under all state changes" (Rosenschein, 1985, p. 12). Regardless of how program structures are compiled or transformed by a learning process, the program embodies the designer's ontology. Rosenschein's formal analysis can be contrasted with Brooks' analogous, but ad-hoc constructive approach (functionally-layered, finite-state automata) (Brooks, 1988); Brooks assembles circuits without spelling out his ontological commitments to world objects, machine states, and relations among them.

Knowledge is the capacity to maintain dynamic relationships.
Agre views the ontological descriptions built into his robot as *indexical* and *functional*. That is, descriptions of entities, representations of the world, are inherently a combination of the robot's viewpoint (what it is doing now) and the role of environmental entities in the robot's activity. For example, the term *the-ice-cube-that-the-ice-cube-I-just-kicked-will-collide-with* combines the indexical perspective of the robot's ongoing activity (the ice cube I just kicked) with a functionally-directed visualization (one role of ice cubes is for destroying bees).

Agre demonstrates that an internal representation of the world needn't be global and objective, in the form of a map, but--for controlling robotic movements at least--can be restricted to ontological primitives that relate the robot's perceptions to its activities. There are two more general points here: The claim that representations are *inherently* indexical and functional (that is, a rejection of the correspondence theory of truth, that representations are objectively about the world) and the claim that the robot can get by with mostly local information about the activity around him. Agre is showing us a new way of talking about knowledge base representations, and demonstrating that a different perspective, that of "dynamics" as opposed to "objective description," can be used for constructing an ontology. It is arguable that Agre's programs aren't fundamentally different from conventional AI architectures; the use of hyphenation just makes explicit that internal names and variables are always interpreted from the frame of reference of the agent, relative to its activities. The important claim is metatheoretical: All representations are indexical, functional, and consequently subjective.

Knowledge is attributed by the observer.
Cohen's work nicely articulates the distinction between designer, robot, behavioral dynamics, and observer's perception that Rosenschein, Agre, and Brooks are all wrestling with.

"AARON draws, as the human artist does, in feedback mode. No line is ever fully planned in advanced: it is generated through the process of matching its current state to a desired end state" (Cohen, 1988). "All higher-level decisions are made in terms of the state of the drawing, so that the use and availability of space in particular are highly sensitive to the history of the program's decisions." Notably, AARON's internal, general representation of objects is sparse; it doesn't plan the details of its drawings; and it maintains no "mental photograph" of the drawing it is producing.

There is no grammar of aesthetics; rather 3-d properties, *as attributed by an observer*, emerge from following simple 2-d constraints like "find enough space." The point is made by Agre, in saying that the purpose of the robot's internal representation is "not to express states of affairs, but to maintain causal relationships to them" (p. 190). The internal representations are not in terms of the "state of affairs" perceived by an observer, but the immediate, "ready-at-hand" dynamics of the drawing process (again, the terms are indexical/functional, e.g., "the stick figure I am placing in the garden now is occluded by the object to its left"). The robot's knowledge is not in terms of an objective description of properties of the resultant drawing, rather the ontology supplied by Cohen characterizes the relation between states of the robot (what it is doing now) and how it perceives the environment (the drawing it is making).

SUMMARY AND CONCLUSION: WHO OWNS THE KNOWLEDGE?

The above analyses demonstrate the usefulness of viewing intelligent machine construction (and cognitive modeling in general) as a *design problem*. That is to say, we don't simply ask "What knowledge structures should be placed in the head of the robot?" but rather, "What sensory-state coupling is desired and what machine specification brings this about?" Figure 1 summarizes the elements of this new perspective.

Briefly, the figure illustrates that a machine specification is a representation that derives from the designer's interpretation of the machine's interaction with its environment. No "objective" descriptions are imputed--how the machine's behavior is described is a matter of selective perception, biased by expectations and purposes. The recurrent behavior attributed to the machine by the observer/designer is a matter of how people talk about and make sense of the world. Furthermore, the specification--usually an external representation in the form of equations and networks--is itself prone to reinterpretation: What the specification means (its "semantics") cannot be described once and for all. The validity of the specification lies in the over-all coherence of the designer's goals, the machine's behavior, and what the designer observes.

Cognitive science research has to date not been driven by such metatheoretic analyses. Most researchers have simply assumed that the world can be exhaustively and uniquely described as theories, and that learning itself involves manipulating theories--a correspondence view of reality. But a radically different point of view has played a central role in methodological analyses in fields as diverse as anthropology and physics. For example, one interpretation of Heisenberg's Uncertainty Principle is that theories are true only with respect to a frame of reference. Bohr himself said, "There is no quantum world. There is only an abstract quantum description. It is wrong to think that the task of physics is to find out how nature *is*. Physics only concerns want we can *say* about nature" (quoted in (Gregory, 1988)). AI and cognitive science research has been based on the contrary point of view that theories (representations and language) correspond to a reality *that can be objectively known* and knowledge consists of theories; consequently, alternative design methodologies have rarely entered the discussion.

Let us recapitulate the emerging alternatives approaches to cognitive modeling. In classical planning, epitomized by present-day expert systems, descriptions of regularities an observer will perceive in the robot's interaction with the world are stored in the **robot's memory**, and interpreted as instructions for directing the robot's behavior. Rosenschein breaks with this idea, instead compiling a state-transition machine from a **designer's specification** of the desired coupling between machine and environment. Agre's work reminds us that regardless of what

compilation process is used, a program still embodies a designer's ontological commitments, and these are fruitfully viewed as **indexical and functional with respect to the robot's activity**. As an artist, reflecting on the robot's behavior, Cohen reminds us that this indexical/functional theory is to be contrasted with **an observer's statements** about the robot's behavior. The essential claim is that representations in computer programs are not objective--true because they correspond to the world--but inherently indexical/functional, relationships between the agent and the world that a designer specifies should be maintained. Moving from engineering "knowledge structures" in an agent to designing on the basis of state-sensory coupling constraints is a significant theoretical advance.

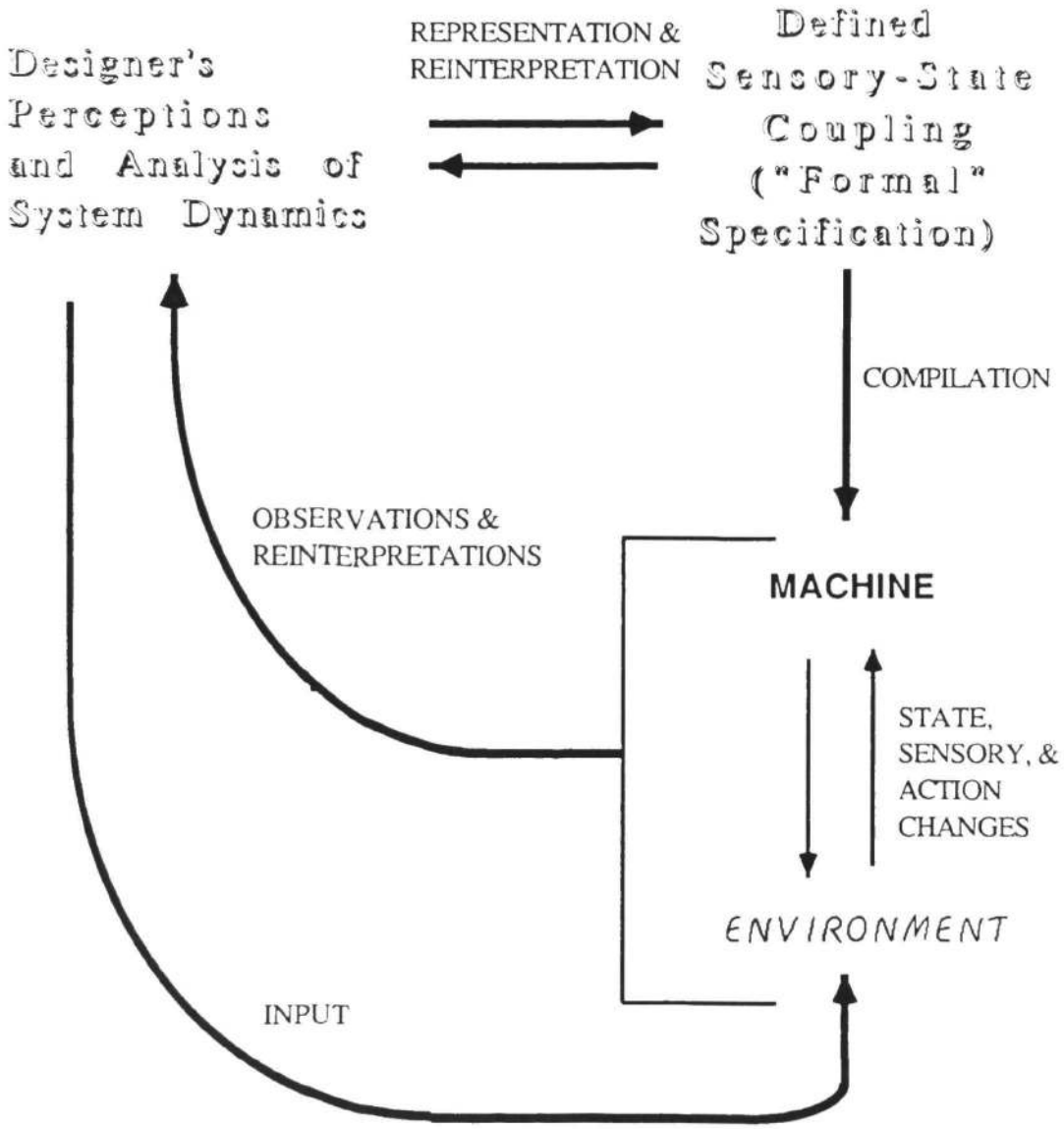

Figure 1. Relation of designer's theory to machine and coupling

However, situated automata research doesn't get to the heart of the matter: Each program still embodies the designer's ontology, which is neither fixed nor objective. Rosenschein in particular continues to speak of an objective physical reality, implying that perception is just a matter of processing data on fixed sensors in axiomatic way (cf. Neisser, 1976). He fails to acknowledge that his coupling specification and background constraints are linguistic entities prone to change under his own interpretation, no less than knowledge structures built into a classical planning system. Formality is not gained by behavioral specification, because these specifications still embody the designer's perceptions of the robot's behavior and his theory of the dynamics of the robot's interactions. Compilation into circuits only changes computational efficiency; the resultant physical structures formally correspond to the designer's original formal notations of "world conditions" and "behavioral correlations." And what these notations mean cannot be objectively specified.

Furthermore, while the robot's structural form is fixed after the design process, the coupling can be modified by human intervention. When a person interprets internal structures during the operation of the program (e.g., providing "input" by responding to the robot's queries), the coupling between robot sensation and action is changed. This interpretation is again an inherently subjective, perceptual process.

Viewing knowledge as relative to an observer/designer's perceptions of dynamic indexical-functional relations between an agent and its environment is indeed a major theoretical reconceptualization of the process of constructing "intelligent agents." However, is a more radical stance possible? Further analysis might focus on the nature of the primitive ontology, specifically to restrict it to sensations inherent in the agent's peripheral sensors (if any) or to primitive perceptual structures that arise in the early developmental interactions of the agent and its environment. From a strict sense, we could claim that the robots described above react to sensors, but never perceive, because they never form new ontologies, new ways of seeing the world. Driving this analysis would be the radical hypothesis that all perceiving is a form of learning and it is dialectically coupled to development of new physical routines. Speaking, for example, is articulating how the world is, conceptualizing, forming perceptions for the first time, not translating internal representations that describe what is about to be said. We must explain how a string like "potentially-attacking-bee" could signify a new way of seeing the world to the robot itself, rather than being a structure that determines its behavior in a fixed, programmatic way. How do we break away from modeling learning by grammatical reshuffling of grammars? For a new beginning, dance, jazz improvisation, drawing, speaking, and ensemble performances of all kinds could be viewed as examples of developing and never fully-definable routines, dialectically coupled to the robot's changing perceptions of own environmental interactions. By this, neural net researchers would move from building in ontologies (however hyphenated or compiled) to finding ways that a process-oriented memory would embody (rather than describe) recurrent interactions the agent has with its world. In short, situated automata research has laid down the gauntlet: How far can we go in removing the observer-designer's commitments from structures built into the machine?

REFERENCES

Allen, C. (1988). Situated Design. Carnegie-Mellon University, Department of Computer Science. Unpublished dissertation for Master of Science in Design Studies.

Agre, P. E. (1988). The Dynamic Structure of Everyday Life. MIT Doctoral Dissertation.

Alexander, J. H., Freiling, M. J., Shulman, S.J., Staley, J.L., Rehfuss, S., & Messick, M. (1986). Knowledge Level Engineering: Ontological Analysis. *Proceedings of the National Conference on Artificial Intelligence*, pps. 963-968.

Braitenberg, V. (1984). Vehicles--Experiments in Synthetic Psychology. Cambridge: The MIT Press.

Brooks, R.A. (1988). *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. (In preparation.)

Clancey, W.J. (1987). Review of Winograd and Flores's "Understanding Computers and Cognition." *Artificial Intelligence*, 31(2), 232-250.

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.

Dietterich, T.G. (1986). Learning at the knowledge level. *Machine Learning* 1(3)287-316.

Dreyfus, H. L. (1972). What Computers Can't Do: A critique of artificial reason. New York: Harper & Row.

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

Kaelbling, L. P. (1988). Goals as parallel program specifications. *Proceedings of the Seventh National Conference on Artificial Intelligence*. Minneapolis-St. Paul, pps. 60-65.

Neisser, U. (1976). Cognition and Reality: Principles and Implications of Cognitive Psychology. New York: W.H. Freeman.

Rommetveit, R. (1987). Meaning, context, an control: Convergent trends and controversial issues in current social-scientific research on human cognition and communication. *Inquiry*, 30:77-79.

Rosenschein, S. J. (1985). *Formal theories of knowledge in AI and robotics*. SRI Technical Note 362.

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

Winograd, T. & Flores, F. (1986). Understanding Computers and Cognition: A new foundation for design. Norwood: Ablex.