Machine Learning, 4, 285-291 (1989)

The Knowledge Level Reinterpreted: Modeling How Systems Interact

WILLIAM J. CLANCEY
Institute for Research on Learning, 2550 Hanover Street, Palo Alto, CA 94304

Machine learning will never progress beyond its current state until people realize that knowledge is not a substance that can be stored. Knowledge acquisition, in particular, is a process of developing computer models, often for the first time, not a process of transferring or accessing statements or diagrams that are already written down and filed away in an expert's mind. The "knowledge acquisition bottleneck" is a wrong and misleading metaphor, suggesting that the problem is to squeeze a large amount of already-formed concepts and relations through a narrow communication channel; the metaphor seriously misconstrues the theory formation process of computer modeling. The difficulties of choosing and evaluating knowledge acquisition methods are founded on a number of related misconceptions, clarified as follows: 1) the primary concern of knowledge engineering is modeling systems in the world (not replicating how people think—a matter for psychology); 2) knowledge-level analysis is how observers describe and explain the recurrent behaviors of a situated system, that is, some system interacting with an embedding environment; the knowledge level describes the *product* of an evolving, adaptive interaction between the situated system and its environment, not the internal, physical processes of an isolated system; 3) modeling intelligent behavior is fraught with frame-of-reference confusions, requiring that we tease apart the roles and points of view of the human expert, the mechanical devices he interacts with, the social and physical environment, and the observer-theoretician (with his own interacting suite of recording devices, representations, and purposes). The challenge to knowledge acquisition today is to clarify what we are doing (computer modeling), clarify the difficult problems (the nature of knowledge and representations), and reformulate our research program accordingly.

1. Qualitative Process Modeling

In the past decade, we have studied knowledge bases and abstracted their designs, so we can describe what we are doing and devise methods to do it more clearly, reliably, and efficiently. Second generation expert systems separate out and make explicit the two processes that are modeled in every expert system: a model of some system in the world (the domain, e.g., a model of an electronic circuit) and a model of reasoning processes (the inference procedure, e.g., a diagnostic procedure) [Clancey 1983]. These two aspects of expert systems are reflected in two dominant, interacting areas of research, called *qualitative reasoning* and *generic systems*, respectively. The focus of qualitative reasoning is to develop notations and calculi for modeling processes in the world. The focus of generic systems

is to develop task-specific representations and inference procedures (e.g., specific to diagnosis, configuration, scheduling, auditing, and control) [Clancey 1985]. These complementary areas of research are integrated in expert systems and associated tools with enhanced capability for knowledge acquisition and explanation. Second generation expert system techniques provide a growing library of abstractions, enabling new programs to be constructed by reusing and refining existing representations and inference procedures. The papers in this special issue make contributions to this research.

Progress to date has followed from the realization that improved expert system explanation, knowledge acquisition, and maintenance depend on abstracted descriptions of the content of knowledge bases and, only secondarily, on the development of alternative representational notations. We call this content analysis "knowledge-level analysis," and contrast it with earlier emphasis on implementation-level concerns (e.g., using rules versus frames). The earlier questions about notations do not go away, but rather are recast in categorical analysis of the nature of the task and system being modeled (e.g., an isolated, designed device versus a biological, open system), how processes are modeled (e.g., classification versus simulation), the inference method for constructing a situation-specific model (e.g., contrasting alternatives on a blackboard versus depth-first, incremental refinement), and the macro structure of the relational network used for describing the domain and inferential processes (i.e., hierarchies, state-transition networks, and compositions of these) [Clancey 1989]. Questions of computer encoding are thus reformulated in terms of process modeling methods that emphasize decomposition and layering of representations.

In short, the first step in clarifying the nature of knowledge engineering is realizing that all knowledge bases contain models of systems in the world and that the expert serves as informant about how such systems tend to behave, how they can be designed or controlled to generate desirable behaviors, and how they can be assembled or repaired. An immediate, important consequence of this realization is that an expert system's performance can be evaluated in terms of the suitability of the model it constructs for the purpose at hand. For example, for medical diagnosis we need to look beyond the name of the diseases output by the program to determine whether the preferred diagnosis covers the symptoms that require explanation [Clancey 1986]. Previously, such consideration of completeness and consistency was reserved for programs using simulation or so-called model-based reasoning. Now we realize that all expert systems are carrying out a modeling task and can be evaluated on this basis.

To spell this out more explicitly, we now realize that qualitative reasoning embraces modeling based on classifications (e.g., a taxonomy of disease processes), as well as modeling based on simulations (e.g., a behavioral simulation in the form of a causal network relating abnormal substances and processes internal to the system being modeled). For this reason, from the second generation viewpoint, we define knowledge engineering as a *methodology for modeling processes qualitatively*, in the form of relational networks describing causal, temporal, and spatial relations. Naturally, it is useful to integrate qualitative with numeric models, and we are belatedly discovering that many expert systems have done this all along (e.g., SOPHIE used qualitative modeling to control and interpret a FORTRAN simulation of its electronic circuit [Brown, et al. 1982]; SACON used simplified numeric equations to estimate stress and deflection, which were then abstracted to select programs that provide more detailed analysis [Bennett, et al. 1978]).

The knowledge engineering community's disparagement of classification goes beyond the suggestion that it is not modeling. Many papers in the literature suggest that classification models are inferior to simulation models and can be entirely reduced to or compiled from them. According to this point of view, physicians talk in terms of syndromes and disease classifications because they do not understand the causal mechanisms causing these processes. A "real" model would reduce disease descriptions to descriptions of physical structure and function. For the most part, this belief is false and belies a fundamental misunderstanding about the nature of system modeling and, more generally, how systems interact.

Disease descriptions characterize *the result of recurrent interaction* between an individual person and his environment. Consider for example tennis elbow. This syndrome cannot be causally explained in terms of processes lying exclusively within the person or within the environment. Rather it is a result of a pattern of interaction between the person and environment over time. As for any emergent effect, it cannot be predicted, explained, or controlled by treating the person in isolation or even by studying the person-environment system over short periods. It is a developmental effect, an adaptation in the person that reflects the history of his behavior. The same claim can be made about the entire taxonomy of medical diseases—trauma, toxicity, infection, neoplasms, and congenital disorders—they are all descriptions of bodily processes after a history of recurrent interactions. Similar examples can be drawn from computer system failures; faults cannot be reduced to changes in a blueprint, but are in fact constantly introduced and prone to change in an open environment. A favorite story at Stanford's SUMEX-AIM is how system crashes were caused every fall when the first October rains wet the phone lines going to Santa Cruz, swamping the computer with spurious control-Cs attempting to get its attention. Such problems aren't fixed by swapping boards.

The consequences of this systems modeling perspective are more staggering than you might first imagine. We are led to realize that beyond the blueprints and functional diagrams of a device being modeled (including the human body), if the device is situated, that is, interacting with an open environment, then a classification model is necessary in order to characterize how the device will appear after it has adapted to a history of interactions with its environment. Such descriptions are necessary in order to describe the state of the device, to explain—historically—how it got into this state, and thus to provide a basis for modifying or controlling the system in some desired way (e.g., to prevent the tennis elbow from recurring). Such descriptions are relative to an observer's point of view; they are not to be confused with the internal mechanisms in the device that produce its moment-by-moment behaviors. To put it simply, a category jump has been made: The system we are now describing is the environment and the embedded device interacting over time, not the device in isolation. Thus, classification models constitute a level of system description, but they cannot be reduced to or mapped onto physical structures in individual devices. As we move from blueprint-like structure-function models, we move from the domain of an isolated system to social, interactive, emergent processes. As Ryle warned us, we make a category mistake if we try to find the university in the members of colleges, the division in the parade of soldier battalions, or team spirit in specific "cricketing operations" [Ryle 1949]. It is no coincidence that Ryle's examples all contrast social organizations with individuals or aspects viewed in isolation. To suppose that classification models can be reduced to the mechanisms of individual agents is to make a category mistake.

At this point it becomes clear that we have to be much more careful in modeling situated systems. We are interested not only in how a device works internally, but how its behavior develops in different interactional environments. This is precisely the province of the human expert, who can tell us what he has observed from experience, as he has participated in the device's operation. For different purposes, we may find it necessary to get the viewpoint of different observers, yielding not one true model of reality, but descriptions relative to different points of view. A great deal can be said about frame of reference problems as a way of synthesizing recent work on situated automata and situated cognition (see [Clancey, in preparation]).

2. Knowledge and Representations

From here it is a simple step to realize that knowledge-level descriptions of human behavior are also descriptions and explanations that an observer gives of a situated system. Knowledge is something an observer ascribes to a human agent in order to describe and explain recurrent interactions the agent has with his environment. Knowledge-level descriptions cannot be reduced to mechanisms in the body of individual agents; they are relative to the observer's point of view and characterize the total system of agent plus environment. Furthermore (now taking a much bolder leap), a knowledge-level description, as a representation, must be expressed in some perceived medium; representations are not stored as or translated from internal structures. For example, when we speak, we are not translating internal representations of what our words mean. Representations only exist physically in an observer's statements, drawings, computer programs, silent speech, visualizations, etc.; otherwise, no observation has occurred. Representations of knowledge are always open to interpretation; their meaning is never fixed or defined, but always relative to an observer's frame of reference. Thus, a secondary level of relativity is interposed by the observer of the observer's representations.

To go back a few steps and summarize, we find ourselves almost overwhelmed with reasons for believing that a knowledge base cannot be associated with structures that were already encoded in the head of the expert:

- Knowledge-level descriptions are attributions made by an observer (the knowledge engineer), involving his own selective interactions with the agent (the expert), his own perceptions, and his point of view;
- Knowledge-level descriptions abstract a sequence of behaviors (what the expert does and says in the course of solving a sequence of problem examples), not single, moment-by-moment responses;
- Descriptions of the device being modeled and inferential processes are informed by the expert's observations and problem-solving behavior, but they are not primarily intended to be the expert's "mental models" or psychological explanations of his behavior.
- To the extent that the processes people follow in gathering data to solve a problem and taking action in the world are intended to be simulated by the expert system, these descriptions characterize a social system (how the expert interacts with his environment), not processes within an individual agent;

- Knowledge-level descriptions have an open interpretation, dependent on the point of view of the observer of the representation;
- Knowledge-level descriptions are always expressed in perceptual space, that is, they are themselves perceivable;
- The human expert, despite often being a theoretician of his own behavior, has no such notations; designing knowledge representations is the province of the knowledge engineer and AI researcher.

Perhaps now we can understand some of Newell's surprising comments about the knowledge level [Newell 1982] (*with my reinterpretations*):

> The knowledge level is not realized as a state-like physical structure, "running counter to the common feature at all levels of a passive medium" (p. 105). *A knowledge-level description is an observer's description and explanation for how a situated system interacts with its environment; it does not correspond to physical structures stored and manipulated inside isolated agents.*
>
> Knowledge can only be "imagined as the result of interpretive processes operating on symbolic expressions" (p. 105). *Knowledge is generated by the process of commenting on (representing) the meaning of perceived ("symbol level") structures.*
>
> "It seems preferable to avoid calling the body of knowledge a memory" (p. 101). "The total system (i.e., the dyad of the observing and the observed agents) runs without there being any physical structure that is the knowledge" (p. 107). *Intelligent behavior isn't physically produced from internal, hidden knowledge representations; it creates them out where they can be seen or heard, interprets them, and is organized by this process.*
>
> "Knowledge of the world cannot be captured in a finite structure" (p. 107). "Knowledge can only be created dynamically in time" (p. 108). *Knowledge is generated by an observer, relative to his point of view, in the process of making sense (modeling).*
>
> "One way of viewing the knowledge level is as the attempt to build as good a model of an agent's behavior as possible based on information external to the agent" (p. 109). *The knowledge engineer's knowledge-level description of the expert emphasizes the expert's awareness and use of materials and circumstances in the environment; that is, it accounts for behavior in terms of interaction between agent and environment.*

As Newell says, knowledge can be represented, but it is never actually in hand. Each statement by the observer captures what he needs to say at any point in time, and each such statement is later interpretable in different ways. Subtle distinctions beyond the scope of this essay are required to further sort out Newell's statements. In particular, when the observer is describing an intelligent agent, a distinction needs to be drawn between knowledge as a capacity ascribed to the agent (dynamically changing through interaction with the environment, not necessarily existing as physical representations for the agent himself) and the observer's representations of this knowledge (perceivable structures, open for interpretation). Hence, we may be ready to return to and build upon Ryle's famous distinction between knowing how (a capacity to perform some action) and knowing that (a representation), in which the capacity to perform cannot be reduced to knowledge-level descriptions about it.

3. Reformulated Research Program

What are the implications for machine learning and knowledge acquisition if knowledge cannot be stored? First, we must adopt a different way of talking about our programs. As I have outlined above, the terms "model-based" and "qualitative reasoning" have been too restrictively applied to qualitative simulation. Adopting the systems-modeling perspective suggests that other approaches should be freely integrated (e.g., linear programming, Bayesian statistics), for we seek whatever models are useful for the task at hand. We are not modeling structures in the expert's head, though we will certainly continue to pay close attention to how he talks and what representations he uses (e.g., diagrams, notational shorthand, calculi). Most importantly, methods from cybernetics, general systems theory, and chaos theory for modeling situated systems need to be incorporated. For the most part, the knowledge engineering community has completely misconstrued the nature of classification and statistical models.

Second, researchers should commit to either providing practical knowledge acquisition tools or studying the nature of intelligence. Providing tools requires more careful attention to the social setting in which expert systems are used; this follows as a generalization and reapplication of the systems analysis given here, focusing on how teams of people interact to solve problems and how job aids can facilitate this interaction. Studying the nature of intelligence will surely continue to involve knowledge-level analyses, for this is the leverage that cognitive science provides over neurobiology. However, a clear separation should be made between knowledge-level descriptions and physical mechanisms. The idea that human-like intelligent behavior could be generated by interpreting stored programs that predescribe the world and ways of behaving must be abandoned, for this view confounds descriptions an observer might make with physical mechanisms inside the agent. Obviously, an agent's own observations, as representations about his situation, purposes, methods, etc., alter his behavior, but these representations are perceivable by the agent himself; they are not stored, matched, retrieved, and refiled by hidden processes [Clancey, in preparation]. Researchers can commit to both knowledge engineering and the study of intelligence, as surely both feed into each other. However, the practical needs of tool users and the difference between knowledge bases and the human mind require a more explicit commitment than before; otherwise, evaluation and choice of methods will be confused.

Finally, the machine learning community in general should attend to the lessons expressed here about how representations relate to human knowledge and make the same commitment required of the knowledge engineers. Much research remains to be done in developing automated methods for improving qualitative models, for example, using explanation-based learning. However, a distinction must be made between the syntactic methods that have been used to date (grammatically shuffling models of processes) and the kind of learning that occurs in the human brain each time a thought is expressed.

References

Bennett, J., Creary, L., Engelmore, R., and Melosh, R. 1978. SACON: A knowledge-based consultant for structural analysis. (STAN-CS-78-699 and HPP Memo 78-23). Stanford University, Stanford, CA.

Brown, J.S., Burton, R.R., and De Kleer, J. 1982. Pedagogical, natural language, and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman and J.S. Brown (Eds.), *Intelligent tutoring systems*. London: Academic Press.

Clancey. W.J. 1983. The advantages of abstract control knowledge in expert system design. *Proceedings of the National Conference on Artificial Intelligence* (pp. 74–78).

Clancey, W.J. 1985. Heuristic classification. *Artificial Intelligence*, 27, 289–350.

Clancey, W.J. 1989. Viewing knowledge bases as qualitative models. *IEEE Expert*, 4, 9–23.

Clancey, W.J. (in preparation). The frame of reference problem in the design of intelligent machines. In K. vanLehn (Ed.), *Architectures for intelligence: The twenty-second Carnegie symosium on cognition*. Hillsdale: Lawrence Erlbaum Associates.

Newell, A. 1982. The knowledge level. *Artificial Intelligence*, 18, 87–127.