

META-KNOWLEDGE AND COGNITION

Avron Barr
Heuristic Programming Project
Computer Science Department
Stanford University
Stanford, CA 94305

In AI knowledge representation schemes, structures that describe other structures are said to represent "meta-knowledge," or knowledge about other knowledge. After describing some studies of human behavior that demonstrate people's ability to reason about what they know and about how they reason, we review the use of explicit meta-knowledge in several recent AI systems. The concept of meta-level knowledge captures intrinsic, commonplace properties of human cognition that are central to an understanding of knowledge and intelligence.

In the last few years, several AI researchers have proposed the use of "meta-level" knowledge representation structures for a variety of tasks. In AI, the phrase "meta-knowledge" generally refers to data-structures in a knowledge representation scheme that "represent" knowledge about other data-structures, as opposed to representing knowledge about "things in the world." For example, a rule in the knowledge base of an expert medical diagnosis system might be annotated with a meta-level description of the rule's history (e.g., which expert entered it or last modified it) or a description of its relation to other rules in the database.

The use of meta-knowledge of this type in AI systems like TEIRESIAS (Davis, 1976) is a key breakthrough in the design of "knowledge-based" intelligent systems. Meta-level knowledge has been used in these systems primarily in the implementation of "introspective" processes: *acquisition* of new knowledge from human experts and *explanation* of the system's reasoning to users. The usefulness of meta-level descriptions for these and other functions has prompted proposals for their incorporation in several new general-purpose representation schemes, like KRL, as described below.

But there is more to meta-knowledge than its typical characterization in AI captures. In human experience, meta-level knowledge and reasoning are an integral part of common, everyday cognitive activity. For example, consider the well-known "tip-of-the-tongue" phenomenon:

You run into someone you have met once before, and you can't remember his name. You remember very well your first meeting at a New Year's Eve party in Oakland, and that he is the brother-in-law of your wife's boss. Then you remember that he has a foreign-sounding name. It rhymes with spaghetti...

You could use all of this knowledge in trying to recollect his name. You would certainly say that you "know his name," even though you can't recall it: You "know that you know it" -- knowledge *about* what you know. It is this *intuitive* knowledge about what we know, and also about how we use what we know, that is the most compelling reason for viewing meta-knowledge as having a central role in human cognition. After exploring some psychological studies which indicate the nature and extent of the role of

meta-knowledge in human memory, reasoning, and understanding, we will examine the recent use of explicit representations of meta-knowledge in several recent AI systems. We will argue, in conclusion, that the apparent difference in character between human meta-cognitive activity and the use of meta-level representations in AI is an important indication of the difference between "representation" and "knowledge" that should be explored further.

Meta-knowledge in Human Cognition

The psychological studies reviewed here deal with human memory, plausible reasoning, and cognitive development. The conceptual framework offered by "meta-knowledge" is essential to understanding these results: It will be argued that much remembering and reasoning is best described as meta-level activity, that at the core of these mental processes people use knowledge about their own cognitive ability, style and experience, and about the extent, origin, and certainty of their knowledge.

It is important to keep in mind that the cognitive behaviors described here are not the results of trick questions or contrived experimental situations. The phenomena that are described are an intrinsic part of human cognition, from remembering to everyday inference making.

Meta-memory: Knowing What You Know

The experience of meta-knowledge by humans is addressed directly in a paper by Kolers and Paley (1976). They point out that the "knowing not" phenomenon is a very common characteristic of human cognition; this is simply where people often know rapidly and reliably that they *do not know* something. Furthermore, these researchers point out that this trait is not easily captured by current "searching" models of memory.

Meta-knowledge and Inference

Allan Collins and his colleagues have for some time been studying "reasoning from incomplete knowledge," that is, what one can conclude from the fact that one doesn't know something. Since the prerequisite to this kind of reasoning is *awareness of not knowing some fact*, these inferences relate directly to the "knowing not" studies, and to meta-knowledge. For example, in the "lack of

knowledge" inference, the fact that you would know some fact if it were true, but you don't remember it, leads you to believe that it just isn't true. Gentner and Collins (1978) suggest two factors used in making the lack-of-knowledge inference. First, one estimates the (relative) "importance" of the fact; the more important it is, the more the fact that you don't remember it implies that it "ain't so." Second, one's own expertise in the topic area is estimated--the more one knows about the area, the more likely not remembering something implies it isn't true. Collins (1978) stresses that much of human plausible reasoning is based on meta-level reasoning about what one knows, and what one would know if some fact were true.

The Development of Meta-cognition

John Flavell has been studying for some time the way children develop increasingly accurate "feelings" about their performance on cognitive tasks involving learning, remembering, and understanding. In a recent paper, Flavell (1979) discusses these results in terms of *metacognitive knowledge* and *metacognitive experience*. Metacognitive knowledge is knowledge about people as cognitive systems, about the cognitive tasks they face, and about the strategies they employ to accomplish them. Metacognitive experiences are the realizations about some aspect of a cognitive enterprise, particularly how well it's going. For example, a person may "feel" that he has memorized a list completely, or that he doesn't understand some instructions.

The point of these studies for the current discussion is that they indicate that the knowledge we have about our knowledge and memory is not simply that we "know that we know X" or that we "believe that adults know what 2+2 is." The meta-level knowledge that appears to be useful to cognitive processes like learning, remembering and understanding, or what Flavell (1979) calls generally *metacognition*, covers the full range of "knowing" about other knowledge.

The Phenomenology of Meta-knowledge

The psychological phenomena reviewed here illustrate the pervasive role of meta-level activity in human cognition. The studies by Kolers and Paley deal with some fundamental properties of human remembering, namely, that one often seems to "know" rapidly and reliably that one doesn't know something, without going through any sort of "memory-scanning" process. In other words, people seem to have "at their fingertips" an idea of the *extent* of their knowledge in some kinds of tasks. Recent research on memory has extended this notion, viewing recall as a problem-solving activity that uses knowledge or descriptions of memories just as other problem-solving tasks use domain knowledge (see Williams, 1977, and Norman and Bobrow, 1979).

Collins's studies indicate that this kind of meta-knowledge may be an integral part of much of people's everyday reasoning. Flavell's work shows that meta-cognition develops gradually in children and that certain meta-cognitive tasks, like estimating how hard a problem will be or knowing when one has understood directions, are performed surprisingly poorly by young children. Once again, what is developing here is not the child's knowledge of the world, but his understanding of what he knows and doesn't know, and of his own cognitive performance. What does it mean to "represent" this kind of knowledge?

The Representation of Meta-knowledge

The AI systems reviewed below all allow the *explicit* declaration of meta-level data-structures in their representation schemes. In other words, the representation formalisms allow encoding of data-structures that "describe" other data-structures. The issues discussed here have come up in many, maybe all, AI systems and are relevant to all representation schemes: predicate calculus, production rules, conceptual dependency nets, semantic nets, procedures, frames, etc. The particular systems described here have attempted to use explicitly represented meta-knowledge.

TEIRESIAS

The use of meta-knowledge evolved naturally in systems developed in what might be called the "Transfer of Expertise" paradigm (Barr, Bennett, and Clancey, 1979). Systems like DENDRAL and MYCIN perform a complex task by using a database acquired from the human experts who are good at the task. The need to give these systems meta-knowledge, knowledge about their own structure and about what they know, developed naturally as part of the effort to endow them with some *introspective* capabilities, in particular, facilities for automating the acquisition of new knowledge from humans, for doing automatic bookkeeping on the database, and for explaining the system's decisions and reasoning strategies to humans. A prototype system for incorporating such capabilities into AI programs, called TEIRESIAS, was designed by Randy Davis, and led him directly to the development of techniques for the explicit representation of meta-knowledge (see Davis, 1976, and Davis & Buchanan, 1977).

TEIRESIAS's various meta-knowledge representation structures are all encoded and used differently within the system, each having its own set of data structures and interpreting procedures. However, the important point is that all of these structures arose out of the effort to implement some new, introspective abilities in the system which were needed to facilitate transfer of expertise interactions with humans.

KRL and KRS

The best known of the new frame-oriented representation languages is KRL, being developed at Xerox PARC. The explicit representation of meta-knowledge was already an espoused feature of the first implementation effort, KRL-0 (Bobrow and Winograd, 1977). Each slot of a unit, or frame, could be tagged with certain *features*, selected from a set of predefined meta-level characteristics, which were used in inheritance and matching.

Besides the *feature* tags, use of an entire description to describe another description, i.e., as a meta-description, was proposed in KRL-0, but was not thought out further until work on the KRL-1 implementation. Recent work on KRL has strongly influenced a theory of the formal semantics of representation languages proposed by Brian Smith at MIT (B. Smith, 1978). Smith's formalism, called KRS, includes meta-level descriptions, called *layers*, as one of its basic characterizations of a representation. The most important aspect of Smith's model of meta-level descriptions is that, unlike the ad hoc character of TEIRESIAS's meta-level knowledge, KRS offers a unified conception of the role of meta-level structures in which the various layers share the same syntax and interpreting process.

FOL

Its explicit representation of its own reasoning mechanisms makes the FOL proof-checking system (Weyhrauch, 1979) of interest in this discussion: FOL makes direct use of meta-knowledge in a first-order logic representation scheme, based on the idea of *simulation structures* which are used to establish the semantics of expressions. The key idea is that, since the FOL proof checker is itself a program, composed of data structures, it is the natural simulation structure to "attach" to a theory of language/simulation-structure pairs, a theory of reasoning. In such a theory one could reason about (prove theorems about) any particular language/simulation-structure pair by using general theorems about L/SS pairs, *meta-theorems*. Although FOL is a very powerful proof-constructing/checking system in its own right, it is uniquely of interest in this discussion because of the neat way that meta-level reasoning fits into the formalism.

The State of the Art

Explicit declarations of the form of the system's representation schemes were necessary in TEIRESIAS to implement "introspective" capabilities like *explanation*. The form of the meta-level representations in TEIRESIAS was ad hoc, but their use was clear. On the other hand, the representation of meta-knowledge as *feature tags* in KRL-O was more uniform, but the ideas about how to use this knowledge were incomplete, involving rather vague ideas of inheritance and matching. Both KRS and FOL have elegant ideas about how to represent meta-level knowledge in their representation schemes, but have not actually specified yet how this kind of knowledge is to be used.

Meta-knowledge and Computation

The first observation that must be made is that there is a qualitative difference between human meta-cognitive capabilities, like "knowing not," "meta-memory," and the "lack-of-knowledge inference," and the current uses of meta-level representational structures in AI systems. In particular, viewing meta-knowledge as additional "facts" or "rules" which describe the object-level knowledge does not completely capture its essential characteristics.

Typical (pre-meta-knowledge) AI programs can achieve expert performance in their domain and yet be unable to answer questions like "Why did you do this?" or "How do you know that?"--questions that a human expert would naturally be able to answer. It was an attempt to implement these very abilities in TEIRESIAS that led Randy Davis directly to the use of meta-knowledge. Humans acquire, as a natural, integral part of their development and training, knowledge about their own reasoning processes as well as knowledge about what they know. These psychological studies of meta-cognitive behavior are important because they deal with commonplace human cognitive abilities, like "knowing not" and "meta-memory," that are very difficult to understand in terms of "storage and retrieval" models of memory (see Barr, 1977, and Restle, 1974). This indicates that there are aspects of "knowing" and "remembering" that have so far remained unexplored in AI research--we have only begun to examine the full fabric of behavior that is the reason we ascribe knowledge and intelligence to people.

References

- Barr, A. *Meta-knowledge and Memory*. Unpublished Working Paper HPP-77-37, Computer Science Department, Stanford, November, 1977.
- Barr, A., Bennett, J., & Clancey, W. *Transfer of Expertise: A Theme for AI Research*. Unpublished Working Paper HPP-79-11, Computer Science Department, Stanford, March, 1979.
- Bobrow, D., & Winograd, T. An overview of KRL, a knowledge representation language. *Cognitive Science*, 1977, 1, 1-46.
- Collins, A. Fragments of a Theory of Human Plausible Reasoning. TINLAP-2, 1978, 194-201.
- Davis, R. Applications of meta-level knowledge to the construction, maintenance, and use of large knowledge bases. Memo AIM-283, Stanford AI Lab, 1976.
- Davis, R. and Buchanan, B. *Meta-level knowledge: Overview and Applications*. Proceedings of IJCAI-5, MIT, Cambridge, 1977.
- Flavell, J. Metacognition and Cognitive Monitoring: A New Area for Cognitive-Developmental Inquiry. Unpublished manuscript, Department of Psychology, Stanford University, 1979.
- Gentner, D., and Collins, A. *Knowing about Knowing: Effects of Meta-knowledge on Inference*, Unpublished manuscript, Bolt Beranek and Newman Inc., Cambridge, Mass., 1978.
- Kolers, Paul and Sandra Paley. *Knowing not. Memory and Cognition*, 1976, 4:6, 553-558.
- Norman, D. A., & Bobrow, D. G. Descriptions: An Intermediate Stage in Memory Retrieval. To Appear in *Cognitive Psychology*.
- Restle, F. *Critique of Pure Memory*. In Solso, R. (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium*. Potomac, Maryland: Lawrence Erlbaum Associates, 1974, 203-217.
- Smith, Brian. *Levels, Layers, and Planes: The Framework of a System of Knowledge Representation Semantics*. Unpublished Draft, MIT, Cambridge, February, 1978.
- Weyhrauch, Richard. *Prolegomena to a theory of mechanized formal reasoning*. Unpublished working paper, Stanford AI Lab, January, 1979.
- Williams, M. Some Observations on the process of retrieval from very long term memory. (Doctoral dissertation, Department of Psychology, University of California, San Diego, 1977). *Dissertation Abstracts International*, 1977.