THE MISMATCH BETWEEN HUMAN AND MACHINE KNOWLEDGE

problems in knowledge engineering

Marie José Vlaanderen

Erasmus University Rotterdam *July 1994*

INDEX

****

Thinking reed ...

346. Pensée fait la grandeur de l'homme.
347. L'homme n'est qu'un roseau, le plus faible de la nature; mais c'est un roseau pensant. Il ne faut pas que l'univers entier s'arme pour l'écraser: une vapeur, une goutte d'eau, suffit pour le tuer. Mais, quand l'univers l'écraserait, l'homme serait encore plus noble que ce qui le tue, parce qu'il sait qu'il meurt, et l'avantage que l'univers a sur lui; l'univers n'en sait rien.
Toute notre dignité consiste donc en la pensée.

Pensées. Blaise Pascal (1623-1662).[1]

... and thinking sand

Silicon; Si, the 14th element in the periodic table of elements. The name silicon derives from the Latin silex or silicis, meaning "flint" or "hard stone".
Silicon is one of the most common of all the elements, occurring in the Earth's crust as sand....
A(n)...property resulting from the electronic structure of silicon is that it functions as an intrinsic semiconductor. [2]

The silicon chip controls our daily life. It replaces boring jobs and creates new ones. It is present in most of the devices that accompany us, that make our lives pleasant or watch over our safety. It is doing a lot of thinking that man used to do himself.
After the Stone, the Bronze, and the Iron Age, we are now living in the Sand Age.

****

[1] Blaise Pascal. Pensées. Édition nouvelle revue sur les manuscrits et les meilleurs textes avec un introduction, des notes et un index analytique par Victor Giraud. (Paris 1943)

346. Thinking makes the greatness of man.
347. Man is only a reed, the most fragile in nature; but it is a thinking reed. It is not necessary that the whole world arms itself to crush him: a vapor, a drop of water is sufficient to kill him. But even when the universe would crush him, man would still be more noble than that what kills him, because he knows that he dies and what advantage the universe has over him; the universe does not know anything about it.
All our dignity thus consists in our thinking.

[2] From the New Encyclopaedia Brittanica. 15th edition.

CHAPTER 1 INTRODUCTION

1.1 THE PROBLEM

In a previous paper (Vlaanderen 1990) I have tried to show the problems that are involved in knowledge acquisition. The transfer between human knowledge and knowledge bases for expert systems seems to be a fundamental and also difficult process in knowledge engineering. Feigenbaum (1977) gave his name to this problem: Feigenbaum's bottleneck.

In the process of knowledge acquisition a knowledge engineer performs the task of extracting knowledge from an expert in a certain field. This knowledge is modeled in such forms that it can be implemented in a computer program that serves as an expert system. The principal problem in this enterprise is that on one side the human knowledge of the expert, and on the other side the way the expert system organizes its knowledge do not match. This discrepancy between human and machine knowledge can be broken down into several mismatches. These mismatches between man and machine are quite obvious, but so far little attention has been paid to this subject in the artificial-intelligence (AI) literature.

In this paper the mismatch problem will be examined from various points of view. The participants in this problem are:

- The human beings: the experts and the knowledge engineer. As a simplification, I do not distinguish between them and average people. Only in certain cases I shall specify the experts.
- The knowledge.
- The machine or the computer program.

Philosophy will provide insight in the "knowledge-part" of the problem. The "human-part" can be viewed from theories in cognitive science. The "machine-part" of the mismatch has its basis in knowledge representation and AI techniques. I propose to select a few themes from the previously mentioned disciplines and see how they can illuminate the man-machine relation.

In chapter 2 knowledge and understanding will be examined. Philosophy has always been the key discipline to solve conceptual questions about knowledge. But is philosophy always relevant in this case? Theories of knowledge seem to be insufficient for the answers we need. Probably a selection from knowledge theories must be made so that each theory can contribute in its own way to the role of knowledge in AI. In the last two sections of this chapter I shall elaborate on "understanding" and argue that this is a concept that can be attributed to humans and to programs. Although, an analysis will show that different levels of "understanding" distinguish man from machine.

Chapter 3 deals with knowledge representation in human beings. For this purpose I have chosen mental models. They seem to be workable hypotheses about knowledge, information processing, and language.

Knowledge representation in expert systems is the subject of chapter 4. Several methods of knowledge representation will be reviewed, evolving in present-day methods. I follow the debates about them and mention some "best bet" methods.

After dealing with these subjects there will be ample evidence to show how important the mismatch problem is in the human - machine relation in general and in the knowledge-acquisition process in particular. Therefore, in the final chapter 5, I hope to indicate the channels that might match human knowledge and artificial-knowledge processing. This might also help the knowledge engineer how to widen Feigenbaum's bottleneck.

The remainder of this introductory chapter is spent to the points of view in the debates that are going on among AI workers about several basic questions. I shall introduce the mismatch problem and what it is all about.
Finally, a brief overview of the problems in knowledge acquisition is given.

1.2 POINTS OF VIEW

The issue of points of view and choosing sides is in artificial intelligence probably more important than in other fields of study. It seems sensible to start this paper with some of these views on several basic AI topics. This can be used as reference for further positioning. Yet, we should bear in mind that present-day AI research has moved away from questions about intelligence as reviewed in this section. Research is now focused on special purpose tools like expert systems. The important issues are now the fundamentals of knowledge representation and shells for the building of expert systems in a faster and consistent way. Other topics are the building of a universal knowledge base, like CYC[3], and building systems without representation, like neural networks.

Rich's definition

One of the more interesting definitions of AI I ever came across is as follows:

"AI is the study of how to make computers do things at which, at the moment, people are better." (Rich 1983 p.1)

Although the vagueness of the definition does not leave much room for insight in AI, it gives us a little insight in the position of the workers in AI. First it shows the humbleness of incompetence: doubting if the whole enterprise of AI is worthwhile. It accommodates the usual laymen reactions to AI when they have heard our explanation. But also serious AI critics, like Dreyfus (1979), form their basic criticism on the essential difference between what people can do and what computers ever will do.

Unfortunately, Dreyfus was beaten by a chess-computer program in 1967[4]. Of course this did not count as a success for AI, because Dreyfus could not talk politics, or weather, or anything else with the chess-machine. He only could have done that with a human chessplayer. Yet, he could

[3] See Lenat and Feigenbaum 1991

[4] See Boden 1977, p 436.

have discussed his emotional problems with ELIZA[5], but then, ELIZA does not play chess as well. So computer systems are inadequate substitutes for human beings, for the time being.

In the world of games, computers can do certain things better than most people. In 1979 a backgammon program won from the human backgammon world champion[6]. But also in other fields of expertise remarkable results take place. The program MYCIN[7] that can diagnose and advice treatment for infectious diseases performs better than the average medical practitioner. Other examples are in the area of geology (mining: PROSPECTOR), chemistry (DENDRAL) to name a few "old" systems. Today, hundreds of useful systems - called expert systems - are commercially available. The state of the art in AI shows that when it comes to specific well defined tasks, computer programs can perform faster and more accurate than human beings. They may even perform better than people who are trained for a certain task.

So far the counterexamples that weaken Rich's definition. The examples that will support the definition remain numerous. There are always tasks that people can do better. This is where the second aspect of the definition is clear: workers in AI will always have a challenge to continue and improve their efforts. The restriction in the definition "at the moment" refers to the tasks that still can be done by people. During the course of time there will be more and more things that can be done by computers, and even better than people can do. A relatively young science like AI needs some provocations like this to get itself established and recognized. So far, this definition serves the purpose of challenging. However, within the discipline of AI, it is necessary to describe the goals and limits of the field more precisely. First of all it should be understood that AI is a name for a complex science with different subfields. Robotics, theorem proving, playing games, perception, machine translation, speech recognition etc., all need different approaches and lead to different results. Some techniques are the same, some are different, but all under the roof of the AI house.

"Artificial" and "intelligence"

The term AI has the disadvantage of being anthropomorphic. When stressing "intelligence" AI gives the suggestion that we are dealing with a simulation of human intelligence. Even that we are trying to make duplicates of human brains. I would not recommend to sustain such a point of view. Human intelligence has the connotation of being a very complex instrument. A human being shows his intelligent behavior in his interaction with other (human) beings and the rest of the world. Simultaneously or successively he[8] can solve problems, participate in a discussion of whatever kind, watch a bird flying over and recognize its species; he can hear sounds and draw conclusions about the origin and meaning of the sounds. He can summarize these events, fill in the gaps of information and use it all for a relevant purpose in the future. To this complex and

[5] ELIZA is a very simple, straightforward, program that simulates a sort of Carl Rogers' psychotherapy. See Weizenbaum 1966.

[6] In July 1979 the program BKG 9.8 defeated the world champion backgammon Luigi Villa in Monte Carlo. The score was 7-1. This was the first time that a board or card-game program won from a world champion (see Berliner 1980).

[7] See Shortliffe 1976

[8] "He", "his", and "him" can also be read as "she" and "her".

well adapted behavior an artificial counterpart would be superfluous. If we stress both "artificial" and "intelligence", only those systems would remain that serve as tools that improve the functioning of this behavior in certain details for certain purposes. In this view AI should be regarded as a one (or few) purpose tool, but not as a general replacement. This position makes way to ignore questions like "can computers think" and "do computers have knowledge" as so many books on AI and philosophy find necessary to ask. Moreover, this seems to me a merely linguistic problem.

The scepticism about AI has more to do with what you expect from it. If one sees AI in terms of machines that can compete with human beings there is enough reason to be sceptic. But if one sees AI as a technique that can do certain jobs better than we do, like vacuum cleaners clean much more thoroughly than sweeps and welder robots in car manufacturing are more accurate than human welders, that seems to be a great advantage already. If you speak of AI only in terms of techniques or systems that are used for one or more special purposes and can do certain things in an intelligent way, scepticism is overtaken by the facts. But if we go into questions like "can computers think" and the like we have to be very careful (see the Chinese room in chapter 2).

The imitation fallacy

Another controversy needs mentioning. A number of AI scientists (e.g., Newell and Simon 1963) think that if we have built a computer program that is able to simulate certain intelligent human behavior, we can also make conclusions about the working of the real human brain. It is like the development of submarines that can teach us how fishes swim! There is also another angle to this. Newell writes *"AI is the study of the mechanisms of intelligence and all supporting technologies.....AI is itself part of computer science, which is the study of information and its processing in all its variety."* But he also claims that *"AI provides the theoretical infrastructure for the study of human cognition."* These positions ties AI and human intelligence together without questions asked, and without arguments to prove. When AI is defined in such manner it leaves no room to discriminate between human and artificial intelligence. To me this is obviously not the case.

Boden (1989) distinguishes clearly: The computers used for mental modeling are digital, serial, general-purposes devices, but the brain is an analogue, parallel-processing, special purpose machine. So, how is it possible to sustain Newell's second claim? Perhaps it is still possible to sustain a modest view such as Miller's (1981) who says that AI will prove to be an incomplete theory of cognition, a theory of certain lower level processing operations that require conscious attention only when they fail.

But, the idea that we can learn from AI processes to get a better understanding of psycho-physiological processes is not only a non-purpose of AI but also a very doubtful hypothesis. I will elaborate on this topic in chapter 2.

Weizenbaum (1976) assumes three types of AI modes:

1. Performance mode: reaching the goal, no matter how.
2. Simulation mode: reaching the goal like human intelligence.
3. Theory mode: reaching the goal by elaboration of a theoretical concept.

The performance mode might work, but then it is doubtful if we can talk about Artificial "Intelligence". A chess program that computes all possible modes before it comes up with an answer (suppose it is feasible in limited time and space) cannot be called intelligent, although the results might seem intelligent. It will be different if the program can deduct a strategy from an initial brute force method and use it for the next games.

The simulation mode can be discarded for the reasons mentioned above. Remains the theory mode. Construct a theory about intelligent behavior and see if it works. May be we should say: "...and see if it stands through the Turing test."[9] Although, with the Turing test the risk of simulation is again present. The theories might vary and do not resemble theories about human intelligence. I want to call this the imitation fallacy. This is the case in all the simulation of human intelligence efforts. Numerous tools that ease human actions have no resemblance at all with those actions. And do we call a hammer an artificial fist? Do we call wheels artificial legs? I want to stress that with complete different principles and the theories that accompany them, artifacts can reach the same or better results.

Interaction between man and system

Thus, AI theories should be theories about how to solve problems, make decisions and other things that usually need human intelligence, in an efficient way with an acceptable result. So far this will be a "no matter how" approach. However, I want to plea for an additional condition. During the development of AI systems or expert systems and during the consultation of the finished system there is a lot of interaction between man and system. It is this interaction that will be an essential part of the whole AI enterprise. Therefore it is also essential that human-information processing is compatible with the machine and the other way around. AI theories should incorporate human-information processing together with other methods. In this paper I want to stress this problem in various ways.

Controversies

A basic controversy in AI points of view can be found in Dennett (1986). He points out the contrast between what he calls the High Church Computationalism (HCC) kind of working in AI and cognitive science (CS), mainly located at MIT; and the Zen Holism approach, employed by AI workers West of MIT. Although the geographic boundary is not that straight, the best exponents of the Zen Holism, Winograd, Dreyfus, and Sellars, are working in the West at Berkeley and Stanford. The HCC point of view is that thinking is information processing, which is computation, which is symbol manipulation, which is strictly formal (e.g., Fodor, Pylyshyn). The Zen Holists show a sceptic attitude towards the possibility of AI to start with; and when they do go beyond that attitude, they see AI programs much more extended than only logic programs. They feel that other things are involved in order to make sound AI programs.

Another way to stress different attitudes towards AI are the Scruffies and the Neats (in Sowa 1984). Sowa uses Bundy's categorization of scruffies and neats:

[9] Because the Turing test originated in 1950, it needs reconsidering. See the discussion in SIGART Bulletin 3 (1992) nr 4.

"The neat view of AI assumes that a few elegant principles underlie all the manifestations of human intelligence. Discovery of those principles would provide the magic key to the workings of the mind. The scruffy view is that intelligence is a kludge[10]: people have so many ad hoc approaches to so many different activities that no universal principle can be found." (p. 23)

and:

"The scruffies dismiss symbolic logic as psychologically unrealistic and linguistically worthless. For them, the important topic is common-sense reasoning, and formal systems are misguided or at least irrelevant. The neats, on the other hand, regard the "network hackers" as developing, at best, trivial variants of first-order logic and, at worst, confused notations with no denotation. Both sides are partly correct. The neats are right in criticizing some of the scruffy programs, but they overlook the importance of a smooth mapping to natural language and the heuristic value of schemata." (p. 137/8)

I agree with the Zen Holists and the Scruffies, since they sustain my previous arguments that in AI much more is involved than formal-symbol manipulating. This is actually the topic of this paper.

My definition of AI

Returning to Rich's definition of AI. Of course, there will be numbers of jobs that can be done better by human beings, but for the many jobs that can be done by computers I would like to give my definition of AI:

AI is the theory of making computer programs that perform certain tasks that can be regarded as involving a certain kind of ingenuity and which cope and interact with human intelligence.
These programs have their own status and merits and should not be confused with human intelligence.

This is not a "nice" definition, according to the rules of how a definition should look like. It serves merely as a framework for this paper.

(In the coming chapters I will use terms that elsewhere sometimes are used in different ways. "Programs" are software programs that can run on a computer. "Machines" are software programs and hardware, like digital or parallel computers. "Robots" are machines and devices that can interact within their environment, can move, perceive, etc.)

[10] "Kludge" is a term used in engineering and computer science to describe something that, from a pure (i.e., a-historical), design-oriented viewpoint, looks messy and inefficient. But it gets the job done. And it may even count as an elegant solution once all the constraints (e.g., the available skills and resources) are taken into account. (In Clark 1989 p. 69)

1.3 THE MISMATCH PROBLEM

The methods of knowledge acquisition and the building of expert systems are often presented to be straightforward and easy to apply. But as stated in the next section, in fact many problems rise while working with knowledge acquisition. Often we read about knowledge acquisition as the bottleneck of expert system building.

What is that bottleneck called knowledge acquisition? The metaphor is well-chosen. We can regard human knowledge as a tangled mass of information stored in a not too wide necked bottle. The expert system is a container with compartments of various sorts. In the knowledge-acquisition process the knowledge engineer must try to get hold of the human knowledge and to put it in the appropriate modes at the appropriate spaces in the expert system. With this picture in mind we can see the problems of the knowledge-acquisition process more clearly.

To substantiate these problems I introduce four hypotheses about mismatches between man and machine.

The knowledge mismatch

The most fundamental difficulty seems to be the mismatch between human and machine knowledge. First of all, a discrepancy exists between the organization of knowledge. Human knowledge is stored in a yet unknown way. The most commonly used hypothesis is the idea of mental models (see chapter 3). Using knowledge is for a human not a conscious process of retrieval and applying of facts and rules, but an unconscious involvement with objects and state of affairs (or their descriptions). The analogy between a mental model and the real world state of affairs gives the solution of what kind of knowledge should be applied. In this process of "applying" knowledge there is no need to verbalize the knowledge. When talking about an expert who "knows" about his field, we are talking about something vague. Knowledge is partly shown in the behavior of this expert, but the logic or justification behind his actions is not shown. The applied rules remain unrevealed too. What we want to have is a sort of materialization of knowledge: a representation of knowledge. Only in a representation knowledge is transferable, can be understood, can be manipulated, etc. Verbal representation is a good solution. As a matter of fact it is the most commonly used kind of representation in human communication. But there seems to be an obstruction when it comes to verbalize knowledge in all its details.

The representation mismatch

There is not always need for an expert to verbalize all his knowledge to non-experts. In the exchange of knowledge between experts (and even between skilled and less-skilled experts) the information is partly presupposed. That is the common-sense knowledge about the world but also the general knowledge about the specific domain, such as vocabulary, concepts, general rules. Even in the situation of knowledge exchange in teaching a great part of knowledge is already presupposed.

In the knowledge-acquisition process - when expert knowledge has to be transferred into a system that does not contain any knowledge - the domain expert encounters great difficulty when he has to verbalize his knowledge. The expert system, on the other hand, uses also a system of

knowledge representation. Most commonly used are representation by frames for declarative knowledge and production rules for procedural knowledge. Essential is the formal mode of these forms of representation.

So far we have mentioned human knowledge represented by mental models, by behavior, by verbalizing (in a natural language that is difficult to formalize). At the expert-system side knowledge is represented by frames and production rules written in programming languages. This incompatibility between man and machine can be called the representational mismatch.

The completeness mismatch

Tacit knowledge, the knowledge that cannot be represented at once, causes another difficulty.[11] The human expert is not aware of all the knowledge he possesses. Only when he is explicitly asked about it, or when he is engaged in a particular section of his expertise, he can verbalize this knowledge. The expert system needs all the knowledge involved to solve the problems for which it will be built. This gap between the expert and the system is the completeness mismatch.

The category mismatch

An obvious, but less mentioned difference, is the disparity between human and machine as two totally different systems. We are not talking about aspects such as difference in representation, but the integrated system of man as an information-processing system (e.g., problem solver) versus the expert system as an information-processing system. The overall approach is so different that comparing aspects seems fruitless. Even the above mentioned differences in representation are incongruent: vague, not verbalized versus formalized; incomplete versus complete. These differences are more profound than a difference, e.g., between programming languages, between frames and scripts, or between the knowledge of expert A and expert B in a particular field. This is summarized by saying that the system of human-information processing is of a different category than machine-information processing. We can call this the category mismatch.

These four mismatches return at the other end of the expert system: the interface with the end-user. It is less of an obstacle for human beings to use successfully the output of a computer system. Still some difficulties in its use remain. The better designed expert systems have explanation facilities to assure the user of its reasoning. Although this is not so much a knowledge-acquisition problem as a problem of expert-system designing, it is worth mentioning.

So far we have given the knowledge-acquisition bottleneck a name: the mismatch problem.

From this brief analysis of human versus machine knowledge, we can conclude that during the knowledge-acquisition process we should constantly be aware of the mismatch in approach between man and machine. Actually, this means for the knowledge engineer that a) he should have a good understanding of human-knowledge processing, and b) he must change, or at least adjust, machine-knowledge representation, so that human knowledge can be better incorporated.

[11] Polanyi (1967) deals with tacit knowledge in a philosophical way. Here, tacit knowledge is regarded as a practical problem.

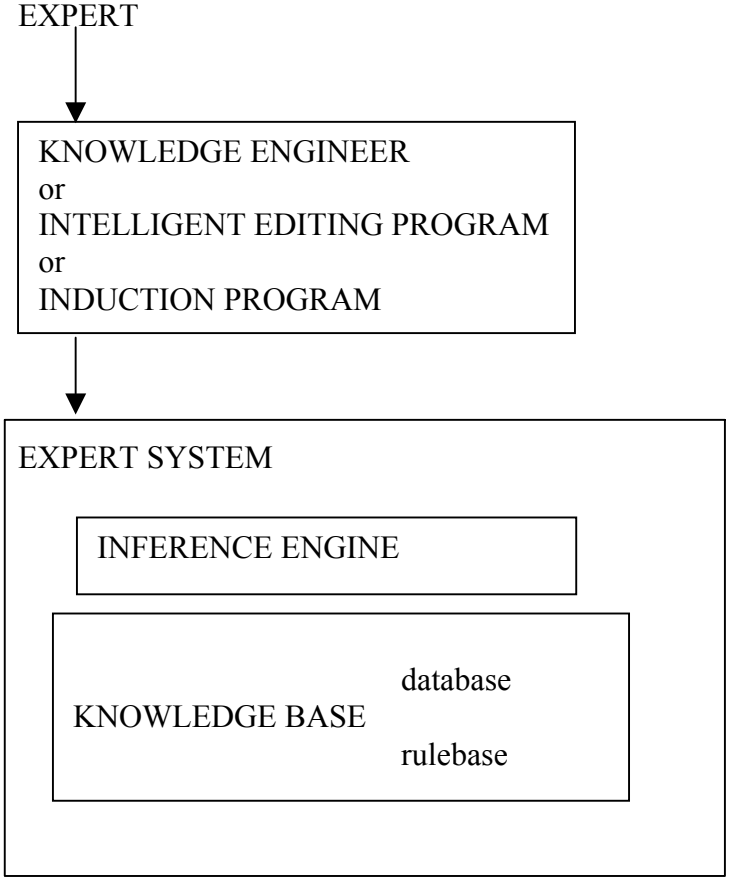
1.4 KNOWLEDGE ACQUISITION

In this section I will give an abstract of the chapter on knowledge acquisition and the building of expert systems from my previous paper to place the topic of the present paper in the right context.

An expert system is a computer program that a) reasons with domain-specific knowledge that is symbolic as well as mathematical; b) uses domain-specific methods that are heuristic (plausible) as well as algorithmic (certain); c) performs as well as specialists in its problem area; d) makes understandable both what it knows and the reasons for its answers; and e) retains flexibility. (Buchanan and Smith 1988)

According to Buchanan et al. (1983) the expert-system-building process has the following outline:

****



********** fig. 1.1 The structure of an expert system-building process **********

It is very important to notice that knowledge acquisition and expert-system building is an interactive process. It would be preferable to draw in figure 1.1 also an arrow from the lower to the upper box.

The sequence of tasks of a knowledge engineer is as follows:
Acquiring a basic knowledge of the domain subject and the goal, or tasks and subtasks, of the expert system. Then, in his encounters with the domain expert he must acquire more precise knowledge, such as the domain concepts. These are the semantic constituents of the domain. Further, the main rules and the subrules, or the strategies of the domain should be made explicit. From here on he can construct a prototype of the expert system. With the prototype the knowledge engineer can expand, deepen and, correct the knowledge with the help of the domain expert. In short, this process is a cyclical process of conceptualization, formalization, implementation, and testing.

In the identification stage the structure of knowledge needs explicitation:
- What class of problems will the expert system be expected to solve.
- How can these problems be characterized or defined?
- What are important subproblems and partitioning of tasks?
- What are the data?
- What are important terms and their interrelations?
- What does a solution look like and what concepts are used in it?
- What aspects of human expertise are essential in solving these problems?
- What is the nature and extent of "relevant knowledge" that underlies the human solutions?
- What situations are likely to impede solutions?
- How will these impediments affect an expert system?

In the conceptualization stage the following questions precede the explicit questions about conceptualization:
- What types of data are available?
- What is given and what is inferred?
- Do the subtasks have names?
- Do the strategies have names?
- Are there identifiable partial hypothesis that are commonly used? What are they?
- How are the objects in the domain related?
- Can you diagram a hierarchy and label causal relations, set inclusion, part-whole relations, etc? What does it look like?
- What processes are involved in problem solution?
- What are the constraints on these processes?
- What is the information flow?
- Can you identify and separate the knowledge needed for solving a problem from the knowledge used to justify a solution?

Depending on the goal of the expert system, the concepts, their relations, the rules and strategies, and the total framework of the domain should be elicited.

In the formalization stage concepts, rules, etc., are mapped in a formal framework (still following Buchanan et al. 1983). Three aspects are important. First the hypothesis space: after the concepts are formalized one has to determine how they link to form hypotheses. The second aspect is the uncovering of an underlying model of the process used to generate solutions in the domain. And third, the understanding of the characteristics of the data that helps to understand the structure of the problem space. Then the implication stage follows from the prototyping and the testing, refining and debugging stage. In these stages no new aspects with the domain expert - knowledge engineer interaction will appear.

Summary

To summarize this section, we can expand the conclusion of the previous section:

1. There is a mismatch between the human and machine approach of knowledge,with
2. Knowledge acquisition is not a part, but an integrated aspect of knowledge-base building.
3. There is a discrepancy between knowledge acquisition and knowledge representation.
4. Questions and answers need to be analyzed on a cognitive as well on a linguistic level.
5. How much of the knowledge is still "hidden" or "tacit"?
6. The various levels of knowledge should be monitored.

7. CHAPTER 2 PHILOSOPHY AND AI

2.1 INTRODUCTION

The answers from traditional philosophy of questions about the nature and extension of our knowledge are unsatisfactory so far we need them for artificial-intelligence theories. (Visser 1985)

Philosophical theories of knowledge are sound and well accounted for. But they serve obviously other purposes than knowledge engineering. It is like using ontology as a base for building houses. The civil engineer needs to know other essential facts about "beings", about the world, than philosophical theories can provide. Fortunately, the civil engineer can switch from metaphysics to physics, but to what can the knowledge engineer switch? The best option seems to be cognitive science. Chapter 3 will deal extensively with cognitive science. However, it seems to be a fair trial for certain theories in epistemology to examine them and see if they can give useful hypotheses for the knowledge engineer.

The sequence of philosophical problems we encounter in the following section is

1. the problem of knowledge,
2. language,
3. meaning, and
4. understanding.

Various kinds of philosophical theories like theories about representation and about language can provide various applicable solutions. In sections 2.3 and 2.4 I will elaborate more on the topic of "understanding" and show in which aspects the machine differs from man.

2.2. PHILOSOPHICAL PROBLEMS

The questions and answers in philosophy are very general and abstract. The theory of knowledge is concerned with academic discussions such as whether knowledge is possible; how knowledge claims can be justified; the difference between knowledge and belief, etc. Theory of knowledge is less interested in the practical questions a knowledge engineer will encounter. However, several philosophers have analyzed knowledge in terms which can be used for our purpose.

The following choice of philosophers and themes is based on personal acquaintance and interests. I am well aware of the fact that many other entries could have been chosen for the purpose of this paper.

Pre-Kantian epistemology

Descartes' method of doubt leads him to accept clear and distinct ideas as a priori truths. Furthermore, Descartes takes God's benevolence as guarantee that our senses would not be deceived. Thus the truth of knowledge, even through perception, is fully based on rationality. Spinoza distinguishes three kinds of knowledge: sense experience, reason, and intuition. Unlike

Descartes' philosophy, sense experience which is derived from experience can be false; not so much because we are deceived, but because we are not able to overlook the whole. We can only grasp the nature of a part if we want to grasp the nature of the whole. Reason and intuition must always be true because the mind understands these matters in a pure and necessary way. This idea of rational understanding Spinoza shares with Descartes. Leibniz also emphasizes the distinction between knowledge from senses and knowledge from understanding or reason. Although the basis for these kinds of knowledge are both innate ideas, the difference is caused by a degree of clarity between the ideas.

Opposed to Spinoza's and Leibniz' rationalism stands the British empiricism of Locke, Berkeley, and Hume. They differ in certain aspects but they all reject the principle of innate ideas. All ideas and knowledge thereof are basically derived from sense perception. Innateness would be a superfluous assumption since perception and human reasoning can account for all knowledge.
Locke derives knowledge from two sources: sensation and reflection. Reflection is the case when the mind observes its own operations and thus forms new ideas. With Locke, theory of language becomes part of the theory of knowledge. He sees knowledge as making judgements, like agreeing and disagreeing. These judgements are in the form of mental or verbal propositions. Here, linguistic skill is a complementary part of knowledge.

Kant

With Kant, theory of knowledge turns into something more substantial for knowledge in artificial intelligence. Kant does not advocate the idea of innate ideas in the sense of his predecessors, but assumes an innate ability to organize our knowledge. Kant makes several useful distinctions. He distinguishes the world as it is in itself from the world as it appears to us. The way our senses work and the way our thought is organized determine the way the world appears to us. He introduces the concepts of space and time as transcendental forms that are innate and therefore precede experience. Originally, Kant has developed these ideas to explain our ability to deal with mathematics[12]. In his *Kritik der reinen Vernunft* Kant elaborates on the idea of transcendental concepts to solve the problem how we are able to acquire knowledge. Instead of proposing pure (static) ideas like space and time, he assumes that our mind bases its reasoning on *Urteilsformen*. Since the mind is in fact making connections and making judgments it seems a good idea to presuppose the innate presence of *Urteilsformen* in the mind. These forms function as molds for our thinking. From these forms the categories are named. These categories are similar to the ones Aristotle used for his logic. The difference is that Aristotle took the categories as static headings for a classification of ideas, while Kant takes them as concepts of judgmental acts and therefore as regulative concepts. The next move for Kant was to make sure that we had a reliable guideline for using the judgments forms. This is elaborated in his Urteilskraft. From here on Kant continues his construction in order to prove the possibility of "pure" natural science and even further to metaphysics.[13]

In sum, we can state that the prerequisites for a human being to have knowledge, according to Kant, means:
1. Sense experience given from the external world (*Empfindungen*)

[12] However, his proof for this prerequisite turned out to be unsatisfactory. Specially Gottlob Frege has criticized this point for arithmetic.

[13] See Edwards 1972 Vol 4, p 310

2. Forms, innate in the mind:
   a. Space and time, to order our sense experiences (*Anschauungsformen*)
   b. Categories, to make judgments
3. *Urteilskraft*, to make the right judgments.

Through our consciousness the *Verstand* makes it possible to have experiences, and thus the world as it appears to us is partly a product of the knowing subject.

Knowledge that we acquire by experience is partly a reflection of that experience and partly formed by the way we acquire that knowledge. For Kant, a concept is a rule that makes it possible to make a unity out of separate perceptions.[14]

Wilhelm von Humboldt

Humboldt (1767-1835) is primarily a philosopher of language. He regards language as the faculty by which man is identified as man. His philosophy of language is focused on the idea of *innere Sprachformen* (inner forms of language). This is a certain view of the world that controls the formation of concepts. There is a dependency between word and thought, as they show a particular outlook of the world. So there is more to a language than only representing the truth, but also a means of discovering previous not known truths. The word is not a copy of the object as such, but reflects the soul's image of the object. Thus, different languages have not the same synonyms in their meaning. Each language has its own characteristics. On this claim, Sapir and Whorf will later build their language-relativity theory.

A synthesis of Kant's theory of knowledge and Von Humboldt's theory of language can be found in Cassirer.

Cassirer

Like Kant, Cassirer claims that the objective world is known by a priori principles. Cassirer assumes that these principles are not static, but dynamic or developing. Cassirer holds that the real world is not a unique and homogeneous thing because every organism has its own receptor system of perceiving the world.
Man has an extra link between the world and himself: the essential function of the human consciousness is the ability of symbolic representation. Man is an animal symbolicum. Man does not deal with the things themselves but actually he is dealing with the linguistic forms, mythical symbols, etc., that are located in his consciousness. Here we find the influence of Von Humboldt. The symbols have no existence in the real world. They function as a connection between the perceived signs and their meaning.

The symbolic representation transcends the perception and places it into a context. So the symbols, like Kant's categories, give an extra dimension to the real world and appear to us as a part of it, depending on the kind of symbol. Ordinary language forms the common-sense world, scientific language forms the sciences, mythical symbols form myths and religions.

[14] See Kitcher 1990

This philosophy of symbolic forms leads to an epistemology that is closely related to a phenomenological theory of knowledge: the real world is identical with this apprehensible world, "perception is apprehension". Cassirer (1975) gives the example of two different meanings of the word for moon. In Latin the word for moon is "luna" which means "the glittering", while the Greek use the word "mén", that is "the measurer".

Conceptualization is a special form of symbolic representation. Concepts are formed after people have named individual varieties. Only the significant ones, e.g., the edible fruits and animals, and other useful object were given names, the rest was neglected.
Cassirer does not construct a sound truth criterium. His objective is to construct a theory of knowledge with a critical approach of language as base. The richer, more differentiated a language is, the better is the knowledge. An explanation how and why a language develops as it does, is not given. But sense experiences and language are inseparable in the course of human evolution. Knowledge is an interlacement of an object, its presence, and its representation, grasped in a symbol function. To put it simple: when an object is experienced and has got a name then there is knowledge about this object. However, there is more to it. The basic problem of this philosophy of language lies in the relation between word and object. The object is not directly given by a word, the word is only a symbolic expression. The function of language is not to depict reality, but to represent by symbols. The symbols are to be understood from within, by inner intuition and related to the cultural environment. There is something which cannot be grasped in the relation thing and word due to the phenomenological position Cassirer has taken.

In the third volume of *The Philosophy of Symbolic Forms* Cassirer projects these forms on a higher level to scientific knowledge. The symbolic representation is here pure concept function. It gives an ordering structure in which objects can fit in.

Peirce

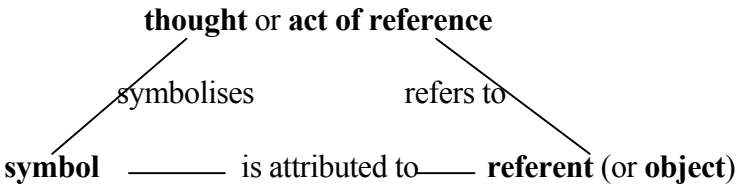
Peirce is well-known for his doctrine of pragmatism. Pragmatism is a theory of meaning. This theory was originally based on his view on human thought and the uses of language, which yields into the concept of "signs". Signs are things that stand for other things. These representations are socially standardized. Thus signs presuppose a society with a system of communication. The pragmatic side of his theory of signs is the practical implication. It should be regarded as a theoretical device - schema or method - to understand the content of concepts. Peirce regards his pragmatic approach as a technique to get better solutions to problems in human communication. Successful communication lies in linguistic and conceptual clarity.[15]

Ogden and Richards - Frege

Peirce's theory was elaborated by Ogden and Richards. In their *Meaning of Meaning* (1972) Ogden and Richards laid out a theory of meaning. The things we sense are sensed as signs. Having sensations is the same as interpreting signs which is done by thinking. This knowledge from the world is the interpreting of signs.

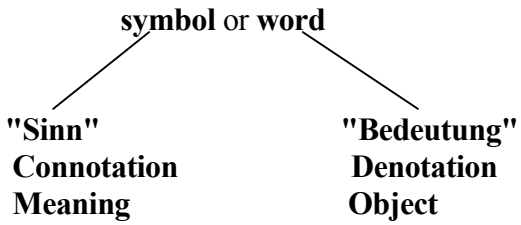
[15] see Edwards 1972 vol 6, p 431

In the triangle



the relation between symbol and referent is a contingent one (except with onomatopoeia). The relation is a social convention within a group of speakers of one language. Therefore it is a loose relation, subject to inaccuracy. Conclusion: the relation between symbol and object is a social one and between reference and referent is not a logical, but a psychological one. And so is meaning.

The triangle above can be compared with another triangle that has been turned over by 120 degrees. It depicts Frege's theory of meaning (Frege 1969):



Frege uses the example of Venus. "Morning star" and "Evening star" have two different meanings (*Sinn*) but only one object (*Bedeutung*). Where Ogden and Richards take thought - the use of a symbol - as starting point, Frege starts with the word. Frege's concern is the logical aspect; Ogden and Richards are pragmatists. Their theory of signs is empirical and behavioristic. Immediate knowledge emerges from sensations. But we sense the objects as signs. Having sensations is the same as interpreting signs which is done by thinking. Thinking is interpreting signs.

The relation between thought and the object is not clear: this might be a cognitive relation. The way a person senses an object constructs a thought in his mind: the art of reference. But in the passage above Ogden and Richards talk about objects as signs. Here they don't clearly distinguish between objects and signs. This is a sort of extra interpretation.

I will try to reformulate this theory in a probably phenomenological way. We assume that objects in the world have an objective, neutral status. Men have sensations of these objects, but since they see them as signs they interpret them in a certain way: the thought. Signs can vary for different persons.
Words are symbols for the interpretation of things and so is knowledge an interpretation of the perceivable world. An object becomes for a person "an object in the way he interprets it" and

symbolizes this interpretation with a symbol or word. Interpretation is person-related and revealed in the use of language.

Recently new approaches to the theory of knowledge have been developed, such as by Rorty. In his *Philosophy and the Mirror of Nature* (Rorty 1979) Rorty shows a sceptic attitude about the foundations of theories of knowledge. Knowledge cannot be "a mirror" of reality. It is determined by personal reflection and social context. Epistemology and philosophy of mind encounter. Man does not have adequate mental means to justify objective knowledge of the world. We are not able to distinguish between what is given from the outside world and what is added by our minds. Knowledge becomes hermeneutics: understanding and interpretation.

Summary

In "good old-fashioned" philosophy we are able to find several aspects that can be used for our purpose to understand human-knowledge processing. Since this is an arbitrary choice it should have no other status than a hypothetical assumption.

1. Both sense experience and "reason" determine human knowledge.

2. Man has the ability to organize his knowledge. He possesses categories that distinguish the various ways in which knowledge is presented.

3. There is a distinction between the world as it is and the way it appears to us.

4. There is an essential interaction between knowledge and language.[16]

5. Man uses symbols to represent the world. Besides symbols there are also thoughts.

6. One object can be regarded in various ways, which can sometimes be reflected in language.

7. Knowledge is interpreting the world, through different descriptions.

8. The problems of interpreting are treated in terms of understanding.

2.3 ABOUT "UNDERSTANDING"

I might have called this section "The Chinese Room revisited". But instead of criticizing Searle's *Gedankenexperiment* I want to use it as a starting point for more *Gedankenexperimente*. They might lead to more insight in the knowledge capacities of programs.

In this section and the following, I want to explain the problems about the use of the term "understanding" [17]. Searle's Chinese room example shows how easy it is to confuse one kind of "understanding" with another. By using other examples I want to show how to distinguish the

[16] These and the following assumptions can also be regarded as expansions of 3.

[17] I shall use quotation marks because "understanding" has different meanings, as will be shown later.

various kinds of "understanding". Moreover, it leads to the conclusion that programs and machines have no "understanding"; and that robots can "understand" in a certain way, although not in the way human beings can.

The Chinese room

Searle (1980) distinguishes strong and weak AI. Strong AI advocates (like Newell) claim that the computer is not only a tool in the study of mind, it is really a mind because programs can be said to understand and have other cognitive states. Weak AI, to which Searle does not object, is merely a strong tool, but not a replica of the mind.

Searle's argument against strong AI is based on the scripts like the ones in Schank and Abelson (1977). A script is a set of possible states of affairs for a certain context. For example, the restaurant script contains the activities around a visit to a restaurant: the sequence of entering the restaurant, being seated and given the menu, ordering, eating, paying. Plus variations: the sequence in a self-serving restaurant, or what happens when the food does not taste well, or when the waiter is rude, etc. The script enables the machine, given a certain restaurant story, to answer questions about what has happened.

In the script-case, the strong AI advocates claim that

1. the machine can literally be said to "understand" the story and therefore can answer questions, and
2. moreover, the machine explains the human ability to "understand" stories.

With the Chinese room *Gedankenexperiment* Searle tries to attack these claims. His starting point is that theories about the mind are best tested if we test it on our own mind. Suppose a person, who does not read or speak Chinese, is locked in a room. He gets a batch of Chinese characters (compare with the restaurant script), then a second batch with Chinese characters and English rules to compare the first and the second batch (compare with the restaurant story and a language parsing program). Finally he receives a third batch with Chinese characters and English rules (compare the restaurant questions and parser). These rules enable him to produce a batch of Chinese characters (compare the answers). Now does this person in the room "understand" Chinese? It seems so to the outside world, but we can be sure that he does not. And so the script-program, the program or the machine for that matter, does not "understand" stories either. All that was going on is manipulation of not-interpreted symbols, in the room, as well in the machine. So the two claims vanish. The machine does not "understand" stories like the man in the room does not "understand" Chinese and since there is no "understanding" there is also no explanation of any human understanding. Moreover in the case of "understanding" an English conversation as a speaker of English, Searle does not think that this kind of human "understanding" follows a similar procedure as the script program.

Many have criticized Searle's attack[18]. I want to elaborate on a fundamental aspect of understanding that Searle does not mention. With better examples the claims of strong AI might be better sustained.

[18] See the comments in The Behavioral and Brain Sciences 3 (1980) and further issues. See Anderson 1989 and Doorman 1989 for methodological critiques.

In the Chinese room only information processing is going on. Input A - program - output B. The operator does not "understand" Chinese, although it seems so. What is wrong? Something is missing. When we compare the data with the semiotic triangle (see 2.2) there are only words or symbols. There are no objects. There are no meanings. Story programs and restaurant scripts manipulate symbols, not meanings. A system that consists of internal symbols, rules and representations, is a system that only computes and compares. The symbols are meaningless to the system.

The Chinese room is about handling syntax (no knowledge) and not about semantics (knowledge). Computer programs are entirely based on syntax, so, by definition, a computer program cannot have knowledge. The machine "understands" the syntax. If there is a frame representation, for example, it also "understands" a semantics related to this representation. But this representation is also formed by symbols. The machine semantics refers to other symbols. Human semantics refers to the real world.

Sloman (1985) explains that strong AI means that the programmed computers can manipulate symbols that they understand. He argues that there is no clear boundary between things that do and things that do not understand symbols. The symbols do not need to be used for external communication. Meaning and understanding are often assumed to be essentially concerned with communication between language users. But representation is prior to communication.

Woods (in Sloman 1983) makes a similar remark: *"By this account, a system that reads stories and answers questions about them, based solely on the manipulation of internal representations with no experiential base, would, from its own perspective, be manipulating meaningless symbols. Similarly a pocket calculator would be a manipulator of symbols whose meanings were externally attributed (even though its internal structure does faithfully model the intended interpretation). On the other hand, a simple robot creature with primitive perceptions can be said to have meanings for its symbols that are not externally imposed if its symbols are tied to those perceptions by some distinguished mechanism equivalent to meaning rules."*

Meaning, as explained in the previous section, comes in when there is an interaction with the world and when this is represented by the system e.g., robots. I come back to robots later.

Rapaport (1988) creates a Korean professor who does not understand English, but is an expert in Shakespeare. He has read everything in translation. Rapaport thinks that the man does "understand" something, although no English. Here we reach a point where two kinds of understanding are existing: "understanding a language" and "understanding Shakespeare". In the last case there is a value added that can be called "cultural" understanding or "aesthetic" understanding. The Korean professor understands to a certain level Shakespeare, but not to a full extent. Understanding a language is not just producing output. It is, in a sense, the ability to cope with the information it carries. That happened with the Korean professor.

Now I am ready to make another *Gedankenexperiment*. I start with a real world situation in which we learn to understand Chinese. Learning is the basic process by which we can understand.

The Chinese restaurant.

Chinese restaurants in this part of the world provide menus for its patrons. These menus are usually lists of courses: a call number, some (2 or 3) Chinese characters, often the Western transcript, so we can pronounce the Chinese, often the translation in our own language, and the price. Suppose we have a menu with only call numbers, Chinese characters and Western transcripts, but no translation. We order the dish by call number and we recall the transcript. The next time we order we pronounce the transcript but get another dish. We were not aware that similar transcripts may refer to different characters, hence a different meal. OK, so we recall the characters as well. After many visits to the restaurant trying out most of the dishes, we are able to make our orders, even by writing them in Chinese characters as well, anticipating certain dishes and, yes, getting the right dish. So we conclude that we understand Chinese, at least the food subset of Chinese. [19]

Conclusion: learning by trial and error in interaction with the world can make us understand a language. In Searle's Chinese room nothing was understood. In Rapaport's Korean room something about Shakespeare was understood, but no English. In our Chinese restaurant we do know the Chinese names for certain dishes given this particular restaurant. We know how to get along with information processing in this place. We might say we "understand" (the food subset of) Chinese.

To compare our experiences in the Chinese restaurant I want to introduce the SHRDLU robot[20]. Winograd's PhD dissertation (1972) is an example of how machines (or systems as he calls them) are said to be able to "understand" language. Like students in biology start working with fruit flies which are rather simple creatures, workers in AI should start with very limited subjects. Winograd took a very small world (the blocks world) as subject for his language-understanding program. (In 1972, AI research was still in its infancy.) SHRDLU is all about language understanding and so a perfect match for the Chinese room. The core of the book is a description of a set of language-understanding programs. To demonstrate this program a blocks world has been created. A robot is able to "understand" instructions like "Pick up a big red block" and then to carry out the instructions. Or it can reply with " I don't understand which big red block you mean" and so on. Winograd's program was never executed with a real robot. The system worked only on a machine. The robot-part was carried out on screen only. However, it is not so difficult to make this program work with a real robot that can see the blocks world with the blocks and pyramids. It can pick up blocks and put them down. The software attached to it can conduct the dialogue. It is obvious that there is a good deal of computation, comparing symbols, and applying rules going on, similar to the Chinese room. But there is also an extra dimension: the blocks world and the robot that is dealing with this world. There are objects, symbols, and the relation between them. There is meaning. Meaning can be defined as a complex relation between objects and their descriptions.

[19] On second thought, the Chinese restaurant is a more complicated example. When we try another restaurant and place our order - that is from a menu without translation - we may get a quite different dish. It turns out that Chinese like to use fancy names for their dishes and what we thought was Chinese for fish and vegetable soup was literary "lotus flower in a fishpond". Another restaurant may give another fancy name for this dish e.g., "green sea with fish". This is rather similar to Frege's "Morning Star" and "Evening Star".

[20] SHRDLU is a meaningless name. It was made up from the 6 most frequently used letters in English texts after AEIO.

Meaning in this sense was lacking in the Chinese room and in the script program. Therefore it was a good example for Searle to attack strong AI. But in the Chinese restaurant case we could relate the Chinese characters to certain dishes, like SHRDLU can relate symbols to blocks.

An objection to the meaning claim might be that a robot does not see objects in the blocks world, but symbols: the zero's and one's on its perception device. What it is doing is matching symbols and applying rules and thus merely computing. If we sustain that claim then people see also symbols. We don't see the bowl of soup but a symbol that stands for a bowl of soup. But of course we don't eat symbols, but soup. So the SHRDLU robot might see symbols, it does not pick up symbols but objects. Objects might be heavy - too heavy to pick up, or made of wet sand and so on. The robot is confronted with an object and its description: not only there is an object, the symbol for it, but also its attributes[21].

Still, I cannot think of a system without a robot or other "real world" device that would be able to deal with meaning in the above sense. And all the knowledge-based systems that are used as expert systems do not do anything more than computing symbols. That is why the weather simulator is not able to cause a rainstorm in the computer room. Ergo: expert systems do not understand what they are talking about!

So it should be clear that programs compute symbols. Robots also compute symbols, but have an extra aspect in meaning. Meaning is a relation between object and symbol.

Knowledge-representation systems that use frames as the meaning of words are described in chapter 4. Still, these programs are manipulating symbols. But we can speak of an internal representation of meaning.

Although Winograd (1972) explains in his introduction to the program that "the system has a simple model of its own mentality it can remember and discuss its plans and actions as well as carrying them out" (p.1), there is not much more than in Schank's scripts. The SHRDLU robot is able to perceive, to learn about the objects it perceives and to apply its knowledge in forthcoming situations. Therefore I would not hesitate to claim that the SHRDLU robot (and not the program) understands the blocks world by itself. There is a relation between (outside) objects and the symbols and a representation.
Yet, there is nothing to sustain, as the strong AI advocates do, that the robot understands in the same way as human beings do.

[21] or "*Art des Gegebenseins*" as Frege calls it

2.4 THE VARIOUS ASPECTS OF "UNDERSTANDING"

In the previous section I have used the term understanding with quotation marks because the appropriate meaning was not always the same. In this section I want to explain these various levels of "understanding".

In the restaurant script (Schank and Abelson 1977) the computer understood^{0} [22] the story, because the answers turned out to be right. Searle's Chinese room analogy showed that this computer does not know English and neither the story involved. The SHRDLU robot understood^{1} English in the same sense as we could get along with Chinese in the Chinese restaurant. The situation was different: there were symbols and real objects. The robot could make connections: the symbols acquired meaning by the presence of objects. Understanding^{1} is involved when we know the relation between one thing and the symbol for which it stands, such that operations on the object can be performed. And also that the attributes of the objects are given in a certain way, to fit into the perspective of the actor.

An expansion of understanding^{1} can be the kind of understanding that occurs in machine-learning programs, in which the relation between object and symbol is grasped. In this situation also a similarity between an object 1 and a slightly different object 2 and their relation with a symbol can be recognized.

A third kind of understanding completes the list: understanding^{2}. This understanding is usually explained in dictionaries as "to perceive and comprehend the nature and significance of". We understand^{2} if we also know the rationale behind the relation between object and symbol.

An example of all the 3 aspects of understanding is knowing to apply $a^2 + b^2 = c^2$ for computing the length of the sides of a rectangular triangle:
A computing program understands^{0} when applying rules upon the given data (symbols).
A robot understands^{1} when relating a particular triangle (with all the imperfections that material things possess) to the formula.
Man will understand^{2} when knowing how and why precisely this formula works.

I have shown that the difference between a robot and a human being is given mainly by the facts that man can understand^{2}, but robots can only understand^{1}. Programs can only understand^{0}.

What is this extra ability of man to understand differently, probably on a different level. This is a difficult question.
Given the literature on this subject, I think that the reasons must be found in the drive of living organisms to survive. On one hand is a teleological aspect: living things are on earth to fulfill an assignment, like producing offspring. In order to do this they need a driving force to meet the requirements. On the other hand is the philosophical aspect of intentionality.

Winograd and Flores (1987) quote Maturana[23]: *"Living systems are cognitive systems, and living, as a process, is a process of cognition."* This organic, or living element, is an extra element that a computer or a robot lacks. Norman (1980) puts the same idea in another way: the

[22] I call this understand^{0} because actually it is not understanding in our common use of the word.

[23] Humberto Maturana. Neurophysiology of cognition. In: P. Garvin, ed. Cognition: a multiple view. (New York 1970)

human body is an organic mechanism in order to protect it, to find food, to deal with its environment.
In other words: living, cognitive, organisms possess an intentionality that makes them live and survive in this point of view.

The ability to understand^2 is a property of human beings like the property of a chemical compound is to form a certain crystal form when under pressure. There are no blue prints of these properties but they can only be perceived when they manifest themselves.

The term intentionality is used in various ways, depending on the context. For my purpose I follow Searle (in Norman 1980). Intentional states are directed at or about objects and states of affairs in the world - states such as beliefs, hopes, fears, desires and intentions. Intentional states consist of a content (e.g., "that you leave the room") and a psychological mode (e.g., "I fear...", "I hope...") corresponding with the kind of act. The content part is not a state of affairs, but a possible state of affairs that only "exists" in the mind of a person. Consider the sentence " John thought that Tom Sawyer is the author of Huckleberry Finn". The whole sentence might be true, but not the content part "Tom Sawyer is the author of Huckleberry Finn", since this does not correspond with a true statement. The psychological part "John thought" is more interesting for a cognitive science point of view. Psychological states can be described, but not fully explained. Like "he hopes …" and "he thinks …", knowing and understanding are also intentional states. These are lacking in the robot. There are only content-statements such as "Mark Twain is the author of Huckleberry Finn" or conditionals like "if the battery is low, get a new one".

In the AI debate one can encounter the question whether we should grant voting rights to the so-called intelligent robots. A sound answer should be: when the robots are aware of their status and ask for voting rights themselves. Then they show that they can understand^2. So the question is whether there is a psychological state in the robot that makes it aware of its battery-dependency, or of its possible voting rights. The answer is straightforward: the robot would not have any urge by itself to care a bit. Only when the conditional "if the battery is low, etc." is implemented and when a device is attached that triggers the conditional when the current gets too low, the robot will react. But the conditional and the device are purely on the content-level. A fortiori, programs will also lack intentional states.

In the last 2 sections I have explicated the different meanings of "understanding". The conclusion that can be drawn from the Chinese room and the other parables is that programs are lacking meaning. The semantic relation between word and real-world object is missing and thus unknown to the program. A program can only operate on the level of symbol manipulation. I have called this "understanding^0". In section 2.3 I have also shown that robots that operate on verbal instructions are able to relate words and objects. One can attribute "understanding^1" to them. Our human way of understanding can be regarded as "understanding^2".

Extrapolating this conclusion to the mismatch problem between man and program, we see that

1. Programs can only understand on a symbol level, i.e., without meaning,
2. Man knows the relation between words and things, and moreover,
3. Human beings can reason and understand on a higher, more abstract level. They also have intentional states.

These arguments that man and programs understand on different levels affirm the claim of category mismatch in section 1.3.

CHAPTER 3

ASPECTS OF HUMAN-INFORMATION PROCESSING

3.1 INTRODUCTION

****

Winograd and Flores (1987) conclude from the writings of Heidegger and Gadamer that we cannot observe the world as it is. We only interpret the world with our prejudices. These prejudices are based in our thinking and our use of language. There is no question of abandoning the prejudices, but at least we can be aware of them.

****

I have stated that the mismatch between human and machine-information processing is the basic problem for a smoothly functioning knowledge acquisition. One solution is to treat human knowledge as machine knowledge, e.g., as proposition logic. But this will give the knowledge engineer a hard time to find all the necessary premises. Moreover, some knowledge will never be revealed. Another solution is to reform the knowledge-representation systems so that they are able to cope with human knowledge. In order to do so we must know much more about the way people deal with knowledge and how they store it.

Cognitive science is particularly apt to reveal how the interpretations we have about the world are formed. It is a recently new expanding field that explores human knowledge. It is an interdisciplinary area between artificial intelligence, psychology, linguistics, and philosophy. Here I refer to cognitive science as the study of human-information processing. Human-information processing deals with the acquisition, representation, and use of knowledge. The human-information processing system is the process of encoding, modifying, and representation (storage) of information received by the various sense organs, the process that we call "thinking", and the verbal and motor outputs based on the previous processes.

It is still a controversial topic how information is obtained and stored in the brain. How are symbol structures represented in the mind and how are operations executed on them. But although theories about human-information processing are hypothetical, it is worthwhile to study them. At least we can try to understand what sort of problems are involved. Two issues are important in cognitive science.

1. The representation: what is the nature of the knowledge structures.
2. The process: what is the nature of the human-information processing system.

Representation and process are probably interwoven parts of human-information processing.

The mental representation depends on the way we perceive and process knowledge. The processing depends on the modes of representation. The hypotheses of human-information processing that are mentioned in the following sections are meant to show its complexity. I think that studying these hypotheses can elucidate our approach to the development of knowledge-based systems and overcome the mismatches.

In this chapter I shall deal with the problems of the constraints on thinking.

3.2 THE PHYSICAL-SYMBOL HYPOTHESIS

A less likely (see chapter 1) hypothesis about human-information processing is the physical-symbol hypothesis of Newell and Simon (1981). Essentially this hypothesis assumes that knowledge consists of symbols of reality and relationships between these symbols; and that intelligence is the appropriate logical manipulation of the symbols and their relations. Newell and Simon developed these ideas in the 1950s as a basis for AI: they assumed then that one would be able to obtain in some way, fundamental atoms of knowledge and the logical relationships between these atoms, and that from these constituents knowledge is reassembled; it has an essentially reductionist strategy. An extension of the physical-symbol hypothesis is the "knowledge principle". That is, the success of an expert system does not depend on the sophistication of its inferencing or reasoning strategy, but on the amount of information it contains on how symbols are interrelated; that is the amount of knowledge it contains (Feigenbaum 1977).

In various publications (e.g., 1978 and 1989) Simon elaborates on the theme that ideas are represented in the mind. He claims that this can be given in an operational meaning and investigated by experiment. He distinguishes in the first place verbal and pictorial representations. Most of the information comes to a person in the form of natural language. If the sentences describe concrete objects they might be encoded as visual images, unlike linguistic strings. But since most information is abstract, it cannot be encoded in pictures. In experiments subjects were frequently unable to distinguish between having seen a particular sentence and having seen a picture whose content was approximately equivalent to that sentence. These experiments might be taken as evidence for the view that there is a single internal representation - a semantic or conceptual representation - into which all inputs, whatever their sensory modality or external coding, are translated.

Simon concludes (1978, p.16):

"I have tried to show that the question, "how are ideas represented in the mind?", can be given a perfectly operational meaning and can be investigated experimentally.....

... The empirical evidence available ... leads me ... to believe that the brain operates basically as a system of labelled associations.
...a plausible guess (how many representations there are): a verbal representation, perhaps not unlike the "deep structures" postulated by the transformational linguists; a visual representation, capable of holding, if not always generating, the information about spatial figures; and a conceptual representation, more "abstract" than the other two and particularly essential for handling abstract meanings.

But whether this is a correct description of the human-representation system is not important. What is important is that we now have a variety of means for exploring representational issues empirically.... Identifying the representation used by the brain is an operational, researchable task."

and still in (1989, p.447):

"The results support the view that pictures and sentences have very similar semantic systems, perhaps the same semantic systems, underlying them."

With the physical-symbol hypothesis in mind it is easy to see the analogy between human-information processing and an operating digital computer. Then it is also possible to attribute intelligence to machines because they operate similarly to human beings. However, studies in knowledge engineering mention that contrary to the expectation of a knowledge engineer, different experts not only talk about a common topic of expertise in quite different terms, but disagree on terminology, without necessarily making this explicit. They use the same terminology in conceptually different ways and would appear to have different knowledge structures. Although these findings might of course be made to fit with the physical-symbol hypothesis, they at least raise the question of knowledge is made up from some sort of absolute primitive elements.

If the physical-symbol hypothesis would have been affirmed then there would be no mismatch problem. However, a number of cognitive scientists (see the next sections) have shown that the physical-symbol hypothesis could not possibly account for the many aspects of human-information processing and thus strengthen the mismatch claims.

3.3 MENTAL MODELS

In an attempt to explain the working of the human-information-processing system, cognitive scientists have introduced the hypothesis of mental models. Among others, Gentner and Stevens (1983) and Johnson-Laird (1983) give bases for mental models and their functions as explanation for the process of human psychology. Mental models are also called conceptual models, schemata, schemes, etc. In general, mental modelers make the assumption that they can explain the working of human thinking to a certain extent. Mental models can be formed about everything of which we have some sort of knowledge, perceptual or verbal. We can say that in mental models knowledge is encoded. A conceptual model is the problem space in which one considers understanding and dealing with the object of knowledge. Interaction of a mental model of the environment with existing mental models of former experiences provides explanation for understanding new situations and how to deal with them. Moreover, conceptual models are formed through analogy with other existing conceptual models and result in identical understanding reactions. (Gentner and Stevens 1983)

Simon (1978) postulates that the brain operates basically as a system of labeled associations. According to him the verbal representations are probably like the deep structures as postulated by the transformational linguists. Visual representations hold or maybe even generate information about spatial figures. Conceptual representations are more abstract and can handle abstract meanings. Mental models are simpler - as models usually are - than the entities they represent. Mental models are analogical representations that are verbal or pictorial and in various degrees of abstraction.

Norman's observations on mental models (Norman 1983) give several details of mental models that make them very limited, also very "human" in their shortcomings, and certainly different

from a clean and smoothly working computing system like Newell and Simon's. In interacting with the environment, people form internal, mental models of themselves and of the things with which they are interacting. These models provide predictive and explanatory power for understanding the interaction. However, as Norman argues,

1. Mental models are incomplete.
2. People's abilities to "run" their models are severely limited.
3. Mental models are unstable: people forget the details of the system they are using, especially when those details (or the whole system) have not been used for some period.
4. Mental models do not have firm boundaries: similar devices and operations get confused with one another.
5. Mental models are "unscientific": people maintain "superstitious" behavior patterns even when they know they are unneeded because they cost little in physical effort and save mental effort.
6. Mental models are parsimonious: Often people perform extra physical operations rather than do the mental planning that would allow them to avoid those actions: they are willing to trade-off extra physical actions for reduced mental complexity. This is especially true where the extra actions allow one simplified rule to apply to a variety of devices, thus minimizing the chances for confusion.

Besides, people often feel uncertain of their own knowledge - even when it is in fact complete and correct and their mental models include statements about the degree of certainty they feel for different aspects of their knowledge. Thus, a person's mental model can include knowledge or beliefs that are thought to be of doubtful validity.
The difference with Newell and Simon could hardly be greater. It is difficult to see how counterparts of human mental models could be designed for computer programs on the base of such analyses.

Johnson-Laird

Johnson-Laird (1983 and 1989b) has a more explicit theory about mental models. He bases the mental-model hypothesis (or theory, as he calls it) upon how people make inferences. His theory has several purposes. First, the mental-model theory can be used to refute the claim that people think in terms of a mental logic. Secondly, the mental-model theory can explain the nature of mental representation and also the relation between language and world. It can show the process by which the meanings of sentences are structured from the meaning of their parts, and on a higher level, how the meaning of sentences is part of a discourse.

Mental models emerge as theoretical entities from man attempts to make sense of inferences, both explicit and implicit. They replace the formal rules of a hypothetical mental logic. Mental models might be artificial, acquired by learning, e.g., mathematical models. Other mental models are natural and acquired without explicit learning, such as language understanding and making inferences. Yet, Johnson-Laird also claims that there are arguments that the construction of mental models, the communication of their contents, and the reasoning on the basis of such representations, are nothing more than computational processes. And this claim is remarkably close to the physical-symbol hypothesis.

Johnson-Laird (1983) has experimented with people making inferences. Not only "abstract data", but also "realistic" knowledge were introduced in their problem task. This can be seen from Johnson-Laird's following descriptions:

"In one realistic condition in this experiment, there was a set of envelopes, including one that was face down and sealed, one that was face down and unsealed, one that was face up with a 50 lire stamp on it, and one that was face up with a 40 lire stamp on it. The subjects were told that there was a general postal regulation: 'If a letter is sealed, then it has a 50 lire stamp on it' and they were asked to imagine that they were sorting letters to make sure they conformed to the regulation. Their task was to determine which envelopes it was necessary to turn over to find out whether or not they violated the rule. In this condition, the subjects, who were English, had no difficulty in appreciating that it was necessary to turn over the envelopes with the 40 lire stamp on it." (p.31/32)

But he also found out that:
" A striking result with one and the same rule is that English subjects over the age of 45, who may remember the postal regulation concerning sealed envelopes, performed well in a recent replication of the envelope experiment, but subjects under the age of 45, who may never have encountered the regulation, performed no better than with abstract material." (p.33)

The conclusion might be that, contrary to what Newell and Simon assumed, people use primarily analogies when dealing with logic problems. Only when they cannot relate the problem with a previous situation they probably tackle the problem as an abstract case. But then they use a kind of reasoning that does not necessarily lead them to the right answer!

According to Johnson-Laird his theory of mental models assumes that comprehension is a two-stage process: first, a superficial linguistic representation of an utterance is set up; secondly, this representation is used in the construction of a mental model of the state of affairs that the discourse describes. In general, a discourse is represented by just a single mental model, which is based on the meaning of the utterances in the discourse, relevant information about the context, and inferences from general knowledge.

At the heart of the theory, in the opinion of Johnson-Laird, lies the following idea: mental models represent the extensions of assertions, i.e., the situations they describe, whereas the superficial linguistic representations, together with the machinery for constructing and revising models, represent the intensions of assertions, i.e. the sets of all possible situations that the assertions could describe. In effect, a mental model is a fragment of many possible worlds: all those possible worlds within which the model can be embedded because the discourse it represents is true in them.
Thus, the structure of mental models corresponds to the structure of a state of affairs. A mental model is therefore different in structure from a semantic network, or a representation in the form of a syntactically structured string of symbols. Both these types of representation have structures quite remote from the states of affairs that they designate.
(Johnson-Laird 1989b)

On one hand, referring to the above described experiment, Johnson-Laird questioned the idea of mental logic: in experiments, it was obvious that people make fallacious inferences. So what sort of logic do they use? How is it formulated in the mind? Is it innate or developed during learning processes? On the other hand, the experiments seems to favor mental models: when individuals have a relevant mental model of a relation, they are more likely to reason about it validly. People follow extra-logical heuristics when they make spontaneous inferences. They appear to be guided by the principle of maintaining the semantic content of the premises but expressing it with greater linguistic economy.

Johnson-Laird gives a sample of mental models as they are supposedly used by people making inferences.

According to Johnson-Laird the sentence
"There are more a's than b's"
is represented as
a---------b
a---------b
a---------b
a

and the syllogism
"All fascists are authoritarians.
Most authoritarians are dogmatic."
is represented as
f = a
f = a = d
f = a = d
f = a = d
a = d
a d

with the suggested - but incorrect - conclusion:
"Most fascists are dogmatic."

Then there are syllogism of the sort "All of the meetings were attended by at least one professor, etc". Euler en Venn are not able to represent these sentences. Also "more than half of the...." cannot be represented this way. And yet people are able to "reason" with these data, although their conclusions are often incorrect from a logical point of view.

Johnson-Laird mentions three kinds of mental representation:

1. The mental models which are structural analogues of the world. They provide a basis for representing premises, and their manipulation makes it possible to reason without logic.
2. The propositional representation which is the mental representation of a verbally expressible proposition, that can be true or false.[24]
3. The images which represent objects. They are analogical so far as structural relations between their parts correspond to the perceptible relations between the parts of the objects represented, taken from a particular point of view.

Johnson-Laird gives three constraints on possible models.

1. Functionalism: The principle of computability: mental models, and the machinery for constructing and interpreting them, are computable.
2. The brain is a finite organism. The principle of finitism: a mental model must be finite in size and cannot directly represent an infinite domain.
3. The principle of constructivism: a mental model is constructed from tokens arranged in a particular structure to represent a state of affairs.

[24] In his article of 1980 Johnson-Laird elaborates on the distinctions between mental models and propositions.

When speculating about how mental models represent the world we should take into consideration that our view of the world is dependent both on the way the world is and on the way we are. All our knowledge of the world depends on our ability to construct models of it. Since this ability is a product of natural selection, our knowledge depends on our biological make-up as well as on the world as it is.

There will be a structural resemblance between the world and our conception because the structure of a spatial model is related to the corresponding physical structure. The trouble is that we have no way of knowing what the structure is (or even of whether the notion makes sense) that is independent from the way in which we conceive the world.

Johnson-Laird holds a functionalist explanation:

1. The mind employs different levels of organization. This principle is easy to discern in language. Inferences from meanings lead to models of discourse.
2. Mental processing at each level takes context into account.
3. Processing at different levels is not autonomous, but interactive.

The three phenomena - the different levels of organization, the dependence on context, and the interaction between levels of processing - all point to one essential principle that mental processes occur in parallel. Higher cognitive processes might depend on other forms of representation.

Critique of mental models

In 1989b Johnson-Laird mentions possible critiques of mental models:
A mental model can be defined as a representation of a body of knowledge - either long-term or short-term - that meets the following conditions:

1. Its structure corresponds to the structure of the situation that it represents.
2. It can consist of elements corresponding only to perceptible entities, in which case it may be realized as an image, perceptual or imaginary. Alternatively it can contain elements corresponding to abstract notions; their significance depends crucially on the procedures for manipulating models.
3. Unlike other proposed forms of representation, it does not contain variables.

Another critique against mental models is that they are an unnecessary explanatory concept. Rips attacks mental models in a significantly titled article "*Mental Muddles*" (1986). He does not think that the mental-model hypothesis adds much to the understanding of human-cognitive processes: Mental models are claimed to mirror directly the structure of the domain they represent, but mental modelers themselves insist that models differ from both propositions and images. Images have less power than has been imputed to mental models. As an example, mental models (conceived as internal simulations) are supposed to be responsible for people's predictions about uncertain singular events. Moreover, mental models are seen as a "semantic" method that is inherently distinct from the "syntactic" method that is more usual in cognitive psychology. This "semantic" approach makes it difficult to find internal models to be fair substitutes. Furthermore, whatever mental models appear to be able to do in the line of explaining psychological semantics can already be done by the old notational systems. If mental models do have an advantage, it is more likely to be in explaining reasoning. But then, manipulation of mental models is not different from manipulation of mental propositions. To the extent that the rules that operate on the models are sensitive to these logical constants, they just

are inference rules. Johnson-Laird's contrast between mental models and mental logic disappears.

My own criticism of mental models is the explicit form in which Johnson-Laird represents them. I was quite surprised by the representations that Johnson-Laird gives of syllogisms and I had a hard time to understand them. I found them highly sophisticated, not less artificial than Euler circles or Venn diagrams. Moreover, Johnson-Laird never gives account how he has obtained these forms of mental representations. And I am certain that only a very few people will use this kind of representation.
On the contrary, Johnson-Laird rules out Venn diagrams and Euler circles as being sophisticated, not natural, mental models. But then, I suppose, everybody can create his own mental models. In a short introspective research I discovered that in cases of easy deductions I use the solution by using common sense. In "Human beings are mortal. Socrates is a human being.", I know how the conclusion will be. For other cases I might use analogies. In more complicated cases I imagine sorts of Euler circles and fill them with A's and B's and so on. I could see in my imagination how the circles overlap or not and could draw the conclusions easily - although not always flawlessly. But then I have had an undergraduate course in logic. I don't think that if your attention has never been drawn to Euler circles or Venn diagrams, you would be thinking of using them.

In the field of (bibliographic) information retrieval an adaption of Venn diagrams is been used to combine key words. "Key word A AND key word B" means that A and B must appear in the same document. While "key word A OR key word B" means that they might appear in different documents. This sounds pretty straightforward. However, this procedure is not recommended, even dissuaded, in systems open to the general public, such as online catalogs of libraries, even in university libraries. Therefore, most library online-catalog systems lead the users through a step by step procedure. First they are asked to enter their first key word. The number of references is shown. Then they are asked if they want to reduce this set by entering another key word. Following this procedure they are not aware that the logical AND operator is used.

Of course, drawing an inference in the form of a syllogism is highly sophisticated itself. People are seldom confronted with such inference problems. Do we ever in daily life have to make inferences of the sort "All artists are beekeepers. Some beekeepers ...etc"? They happen only in undergraduate logic class. By then we are also taught about the mental models that can cope with these syllogism, the Venn diagrams and Euler circles.

Conclusion

Notwithstanding the doubts that Johnson-Laird's inference example raises, the basic idea of mental models, in one form or another, might be a plausible hypothesis about the way human-information processing works. The experiments in which Johnson-Laird speculates less about the exact construction of mental models show evidence that people use schemes or models, rather than separate bits of knowledge, to process information. This was confirmed by experiments done by Holland et al. (1986). Let us therefore conclude that, although we are not sure about the factual outline of mental models, people process information rather with models, schemes, or structures than with primitive symbols.

3.4 CATEGORIES AND CONCEPTS

Another aspect of human-information processing is categorization.
Categorization is a basic cognitive process. Man has learned to understand and to verbalize the objects and states of affairs he perceives. He categorizes while perceiving. Categorization is abstracting into concepts. Abstraction means to select the essentials and leave the fringes. Some information gets lost during categorization. This would not happen with a holistic approach in which all the knowledge components would be given as an integral image. Categorization is simply an economic solution for our cognitive tasks.

What is a category

Smith (1990) takes a category to be a class of objects that we believe belong together. Using the word "believe" suggests that we are dealing with a psychological concept. It turns out that the boundaries of categories are not that sharp and it is sometimes not clear to which category an object belongs. Categories can be made to assign objects with the same properties to one category. Suppose we assign all objects with the same weight and shape to one category. Golf balls and tangerines will then be in one category. The reason we prefer not to do this is because it does not serve any purpose. While objects that are physically not very similar like watermelons and - say - grains of rice are in the same category as food. This makes sense, because we might like to consider the various food products in terms of caloric value. Or we like to put them in the same category because we buy them in the same shop. Categorization is mainly a product of convenience.

There are more pragmatic aspects that make us to use categories.
We use categories as labels or codes for objects we have perceived or otherwise have experienced. It is more economic to remember objects as a member of a category then as an individual. Moreover, it is not only easier to remember something as a category but also to talk about it or to make inferences. If we know certain properties of one member of the category it is likely that other members have the same properties. Considering the conditions that make an object belonging to a certain category, that is, e.g., if we submit watermelon and rice under the category food it is because both are edible and can be digested by people. So if we assign caloric value to watermelon, which is a property of edible things, it is likely that we can do this with rice as well.

There is also the aspect of similarity of objects in the same category. Usually there is a physical similarity, but sometimes not (as with watermelon and rice). Although the similarity helps a great deal to assign an object to a certain category in the first place, this may turn out dangerously. How do we distinguish between edible and poisonous mushrooms? Normally prototypes can be recognized as basic examples for a category. So apple will be a prototype for the category fruit. People tend to need more time to submit an object to a category the more it is defiant of the prototype, like date or raisin to the category fruit; ostrich versus robin to the category bird. On a higher level of cognitive faculties the word category is used for such abstract concepts as space and time. These categories can be regarded as prefab cognitive tools. Categories provide a sort of lattice to select aspects of knowledge.

In the philosophy of psychology categorization is brought forward by Pylyshyn (1978), who argues against mental images as internal representations. He prefers to use cognitive representations that are different from linguistic descriptions as well from images. According to Pylyshyn, the representation is rather selective: after seeing an object or scene we are not able to reproduce all the details. The same is true with reproducing a verbal report. This means that categorization is active. Pylyshyn uses the term structural description. Structural description should consist of concepts (in Kant's sense of categories of understanding). They should bear a referential relation to the objects they represent (rather than resembling). And they are not entirely part of the visual or linguistic information processing.
In Pylyshyn's view it should be possible to explain human-information processing with machine-information processing, but that makes his hypothesis of structural descriptions rather doubtful. However, his other arguments for this kind of mental representation are based on psychological experiments.

A psychological approach is given by Miller (1981), who distinguishes between perceptual presentations and symbolic representations. The level of symbolic representation does not only build on the cognitive categories established at the level of perceptual presentation, but introduces many conventional categories that our ancestors have found useful. The two levels interact closely together. The symbolic representation influences the perceptual presentation in subtle ways - it influences what a person pays attention to and what perceptual distinctions will be drawn and remembered. The symbolic component does not simply label the output of the perceptual analyzer, but it also controls the input to it.

There is no consensus whether categories should be considered as being verbal or pictorial. Haugeland (1981) thinks that the mind is a human-information-processing system that is not related to language. Language is an external mapping, while the mind, or the human-information processing system is a private internal mapping without relation to language. There are reports about children that have been deprived of any kind of language acquisition during their childhood but are able to tell later about their pre-language experiences. More common known are studies in dyslexia about persons who cannot understand words, although they do know them. This should prove that mental representations are not necessarily verbal. But it does not prove that mental representations are necessary non verbal. Further experiments have shown that visualisation helps the understanding of verbal texts. Verbal recall without visualisation is seldom accurate.

Rosch (1978) adds an economic aspect of categories: She distinguishes two principles of categorization. The first is the principle of cognitive economy that implies that with a minimum of cognitive effort the maximum of information can be provided. The other principle is that the perceived world comes as structured information. This structure is already given at the human-perceiver side by culture and language. Our perception is based upon partly what we predict or assume.

Concepts

Categories and concepts are closely related. One might say that a concept is the mental representation of a category.

Arnheim (1969) gives an account of how concepts play a major part in thinking. Thinking deals with objects and events only if they are available to the mind in some way. This can be accomplished by abstraction.
Abstraction, which is mainly the formation of concepts, is a withdrawal from direct experience. This assumes a dichotomy between perceiving and thinking. One sees only particulars, but one thinks in generalities, and therefore, in order to think one must sweep the mind clean of perceptual material. Abstraction is supposed to perform this function.

The study of concepts is central to every discipline of cognitive science. In cognitive psychology and philosophy of mind, concepts are assumed to be the basic constituents of thought and belief. In AI proposals about knowledge representation are proposals about concepts. Concepts promote cognitive economy (hence we can do with much less information). Concepts also allow us to bring to bear our past experience on present concerns. A third function of concepts is that they permit inductive inferences.

Concepts can also be used in definitions or as prototypes: the best examples of a category with typical properties is called a prototype. Under the prototype view an object is categorized as an instance of a concept if it is sufficiently similar to the prototype. The similarity is being determined in part by the number of properties that the object and prototype share.[25]

The similarity-to-prototype approach has been generalized to concepts other than natural kinds, especially artifact concepts like furniture and clothing. The phenomena for natural kinds hold for artifacts as well 1) people do not have fixed definitions of most artifacts; 2) the instances of any artifact concept vary in typicality with typical instances being easier to categorize, remember, name, and reason about; and 3) more typical instances of an artifact concept have non-necessary properties that occur frequently in other instances.

Schemata

The terms "schemata", "mental models", "mental representations", etc., are often interchangeably used. Some authors prefer not to use the word "mental model" at all, in order to avoid confusion with Johnson-Laird's theory of mental models. Or, like most, they use the term that seems to be the most suitable. Some authors give different descriptions for the term schema and the term mental model.

A schema can be seen as a format in which data get their assigned position. This enables us to interpret the data coherently, while irrelevant information will be ignored. In Neisser's (1976) view schemata function as plans for finding out about objects and events to obtain more information. A schema with information can generate upon this information another schema to obtain other information, e.g., increase, specify or alter. By acquiring more and more information the perceptual schemata alter in the same pace. So we never have the same perception more than once: the previous perception has altered the new schema.

When we observe very young infants we must conclude that certain perceptive schemata are already present. When a baby hears a sound it turns his head toward the right direction. He can

[25] Categories can also be used as prototypes. See previous paragraph.

follow objects when moved or reach out to objects. So the suggestion that some schemata are innate is very strong.

Language of thought

Fodor (1975) conceived the idea of lumping categories, concepts, and schemata together in one "*language of thought*". He attempts to construct a mental - but non-linguistic - language, as part of his theory of cognitive processes. In the traditional view, the medium in which we think would be our native language, because that is also the language in which we express our thoughts. However, Fodor rejects this claim because of the existence of nonverbal organisms that are able to think, at least to show cognitive behavior. (Considering action, learning concepts, and integration of perception is also considered as "thinking".) Therefore he postulates that animals in general "contain" representational systems that are already active before a natural language is learned. Learning a language, even the first language, involves learning what the predicates of the language mean, that is to determine the extension of these predicates. Since these operations cannot be done in language (in case when learning a first language) there must be another kind of internal representation or language of thought as Fodor calls it. This cannot be learned, otherwise one falls in the pitfall of infinite regress. The language of thought must be "known", or at least must be present: it is innate. However, the richness of the mental representation and the ease by which human beings can communicate in language makes it very likely that their fully developed language of thought is close to natural language. In general, the internal representation is not a fixed one. It depends on factors such as task, motivation, and environment. If this environment is natural language (versus e.g., pre-language children) then the internal representation is correspondingly. This does not exclude mental images in pictorial form - these will be involved when the relevant information is given in pictures, or otherwise nonverbal.[26]

Concluding remarks

In this section I have elaborated on categories, concepts, and schemata, since they are believed to form basic elements in our cognitive faculties. The different approaches strengthen the idea that we process information in structures rather than in primitive elements, although the state of the art of the "theories" is still far from "explaining" how human beings process information.

[26] In Normore (1990) is explained how Ockham has already followed the same train of thoughts as Fodor about how a mental language is constituted.

3.5 PRACTICAL ASPECTS OF HUMAN-INFORMATION PROCESSING

****

An expert is one who does not have to think he knows
(Frank Lloyd Wright).

An expert is a person who has made all the mistakes which can be made in a very narrow field (Niels Bohr).

****

After having given an overview of approaches in cognitive science, I want to mention some of the problems in the context of practical interaction between human and machine-information processing. With these examples the mismatches will become more explicit.

Human versus machine thinking

Johnson-Laird (1989a) mentions five distinctions between human and computer expertise:

1. Humans can treat errors not merely as cause for excuse but also as the occasion for a revision in their knowledge of theories. Expert systems are normally not confronted with their errors.
2. Humans can do much more with their knowledge than systems can. They can use their knowledge in a flexible way to answer many different sorts of problems.
3. Human experts have a large amount of tacit knowledge that they cannot readily articulate in words.
4. There is a difference in semantics. Humans have knowledge about things, whereas expert systems merely have sets of conditional rules which are denotationally interpretable by the human user, but not by the program.
5. Humans and systems make inferences using rather different mechanisms.

And moreover, humans have mental models.

If we want to make existing systems more like an expert then we have to bring expert and machine knowledge closer together and explore some new AI techniques. Johnson-Laird shows a logical rule that can be expanded by adding an exception. This is the sort of exception that is out of the question in traditional logics, but quite normal in common sense or expert reasoning.

Consider an expert-system rule:
"if p, then r"
In logic no matter what is conjoined to the antecedent the resulting rule will still be true:
"if p & q, then r"
But:
"If infection 1 is x and infection 2 is y, then administer penicillin"
The rule will be rejected in:
"If infection 1 is x and infection 2 is y and the patient is allergic to penicillin then do not administer penicillin, but....."

Now the inconsistency between the first and the second rule is only obvious at a deeper level of knowledge. People are used to crisscross through the various levels of knowledge, without noticing on what level they are. The machine is always at the same level, and if it is able to switch levels there must be a coordinator that keeps track. Therefore, non-monotonic logic would not be able to cope fundamentally with this type of reasoning. Moreover, the human expert has a more flexible way to deal with knowledge. He possesses knowledge, outside the domain that is directly involved, and this allows him to make more easily analogies and inductive reasoning. He possesses common-sense knowledge so he can avoid the traps of logic as in the penicillin case. He possesses a mixture of sensory experience and verbalized matters of fact. He has learned to be adaptive.

In a reaction to McCarthy and Hayes[27] in which they claim that AI programs should be written in a formal language, Sloman (1971) claims that, e.g., analogical reasoning, non-logical reasoning, and the use of non-linguistic representations, are good alternatives for formal languages. Specially when it comes to spatio-temporal environments. Sloman brought intuition, insight, apprehension of relations between universals in against the claim that only valid modes of reasoning are those which use logically valid inference patterns.
Analytical philosophers would always work with the second, but they cannot account for intuition etc, the way people reach their conclusions. Sloman argues that true or false statements not only depend on their structure, but also on the facts. He also claims that changes in the configuration, represented in an analogical representation, can be indicated simply by moving a dot whereas very many changes in linguistic descriptions of relationships would be required. In sum: linguistic representation is not always the best representation.[28]

Expert thinking and the knowledge engineer

This and the following paragraphs are more explicitly about the cognitive processes in expert thinking.

Woodward (1992) gives an overview of the knowledge-engineering process, using the KADS[29] methodology. The knowledge is acquired, transferred, captured, and modeled. This emphasis directs our attention away from the processes that produce, organize, and represent the knowledge: away from notions of learning, comprehending, and communicating. It is important to incorporate an understanding of cognitive processes in the development of theories of knowledge acquisition. The epistemic concepts must be clearly defined, the epistemic concepts must have a base in psychological processes and the representational formalism must reflect the intended epistemic and psychological meaning. Therefore, Woodward sees knowledge engineering as a set of cognitive tasks rather than as a set of activities which acquires and models knowledge structures in one form or another. Summarizing:

[27] Some philosophical problems from the standpoint of AI. In: M.L.Ginsberg, ed. Readings in nonmonotonic reasoning. (Kaufmann 1987)

[28] See also Haugeland's pictorial frames.

[29] See for KADS, e.g., S.A.Hayward, B.J.Wielinga, and J.A.Breuker. Structured analysis of knowledge. International Journal of Man-Machine Studies 26 (1987)

1. Knowledge engineering represents an ordered collection of cognitive activities.
2. The purpose of knowledge engineering is to use cognitive processes to produce a model. Emphasis is placed on the information-processing activities which lead to a final model rather than on the characteristics and content of the final model.
3. Knowledge engineering is done in a cognitive environment.
4. The cognitive environment is characterized by sites and levels of cognitive activity.
5. An understanding of the cognitive environment acts as a basis for developing tools, techniques, and procedures to support and/or replace these processes.

Rather than emphasize knowledge support a complementary view of cognition support as a basis for tool, technique, and procedure development is suggested.

From his experiments with mental models, Johnson-Laird (1983) found data about the difference between conscious and unconscious knowledge. He thinks that the demarcation between what is accessible and what is inaccessible to consciousness is related to the difference between knowing that something is the case and knowing how to do something (as in Ryle). Knowing that p, implies that one has conscious access to the proposition that p is true. Knowing how to do p may depend on conscious access to some of the parameters controlling performance, but it also depends on low-level processors with internal operations that cannot be inspected by the operating system. To try to know (by the knowledge engineer) precisely what one is doing can be very difficult.

Bechtel and Abrahamsen (1990) show us another approach to human cognition. The way we use categories determines also other cognitive activities. The way we deal with categories makes it not likely that our knowledge is memorized as propositions, but more as a framework of knowledge. That is also the way we perceive. The authors mention studies showing that when learning skills, students start with learning rules (in a propositional form) but once they have acquired expertise they recognize the situation and the action acquired as a pattern. The problem is that as soon knowledge must be communicated (which usually happens in language) it happens in propositions. But the experts have already left this "propositional" stage behind! This was the reason for Charniak & McDermott (1985) to consider the problems of the translation of mental images of knowledge into a verbal representation. They focus that experts recall usually cases, not rules, so the knowledge engineer should ask them about cases.

Knowledge in context

Compton and Jansen (1990) look at the relativity of knowledge. If all knowledge is only true in a context, then all knowledge is relative. It only exists in relation to other knowledge and there is no absolute underlying knowledge on which the rest of knowledge is built. So much is clear from twentieth-century discussions in the theory of knowledge, that it is not likely that the underlying primitive knowledge is made up from sense data. On the contrary, in normal cases what we perceive depends on the contexts in which the perception takes place. This is for the building of knowledge-based systems a great problem.

But even when it would be possible to specify underlying knowledge, the experts will always give knowledge in context, with rules that conflict, subsume, overlap etc. If the knowledge engineer insists, the experts can express knowledge that is more complete, consistent and

coherent. Although the danger might be that they are constructing something, making something up to satisfy the knowledge engineer. It is even suggested that the knowledge that experts provide is essentially a justification of why they are right, not the reasons why they reach this right conclusion.

Cognitive emulation

More about human-expert thinking is given by Slatter (1987):
Expert performance depends on large amounts of domain knowledge.
Experts know when a problem is outside their area of competence, or when to break general rules in order to handle exceptions.
Experts can reorganize their knowledge into more appropriate forms. Experts are partly capable of reflecting on their own cognitive processes (meta-cognition), and about their own state of domain knowledge (meta-knowledge). Experts' reasoning and knowledge are frequently inaccessible. Expert cognition lacks both computational and representational power. People's awareness of their own mental processes is rather limited. The proceduralization of knowledge and automatization of cognitive skills, that accompany the development of expertise, serve to make expert thinking even less accessible to introspection. The cognitive correlates of expertise, whether beneficial or otherwise, are essential domain-specific in effect. Thus outside his specialist area any cognitive advantage the expert may have enjoyed inside the domain quickly disappears.

Another underlying theme in the development of expertise is a greater reliance on pattern recognition and memory (stored knowledge) at the expense of deductive reasoning. At another level, though, experts often show an impressive ability to reflect on, and flexibly control, their high-level task strategies. But precisely how automated skills and control strategies combine in expert problem solving remains poorly understood. Tacit knowledge refers to the tacit inferences and global knowledge that provide a non-articulative framework for human reasoning, including expert reasoning. Slatter illustrates this with knowledge-oriented approaches in knowledge acquisition, like e.g., the levels of knowledge acquisition in KADS.

The failure to analyze knowledge at any of these intervening levels - in particular, the epistemological level - may be at the root of several knowledge-engineering problems. From linguistic data to implementation language is too large a gap to have to bridge in a single step; and valuable additional knowledge may be lost by not analyzing at intermediate levels. Epistemological analysis can make explicit types of knowledge that are represented in expert cognition only in a highly proceduralized, highly inaccessible form. That is, for the expert such explicated knowledge may not in fact exist. Nevertheless, when one attempts to build expert systems one might take account of the following
five key issues:

1. The extent to which the standard three-element[30] system architecture of expert systems corresponds to the architecture of human cognition.
2. The psychological plausibility of particular knowledge representation formalisms and systems architectures.
3. The requirements to make expert systems compatible with the limitations of human-information processing.

[30] Static knowledge, dynamic or temporal knowledge, and knowledge to perform inferences.

4. The explicit representation of knowledge.
5. The handling of uncertainty in expert systems inspired by cognitive psychology.

Slatter concludes that, although there is no sound and commonly accepted theory in cognitive psychology, the best way is to adopt an approach based on an information-processing view of human cognition.

The conceptual model

An example of how mental models can explain our dealing with systems, such as a calculator, is given by Norman (1983). He distinguishes:

The target system (= the system that the user is learning): t.
The conceptual model of that target system (invented): C(t).
The user's mental model of that target system (natural): M(t).
The scientist's conceptualization of that mental model: C(M(t)).
(These are often not verbally elicited.)

The conceptual model is taught to the user and must fulfill 3 criteria: Learnability, functionality, and usability.

M(t) reflects his belief about the physical system acquired through observation, instruction, inference, etc. There should be an observable correspondence between the system and C(M(t)). Also the C(M(t)) must have a model of the relevant human-information process and knowledge structures that make it possible to predict and understand the system. As a matter of fact, Norman claims that the major purpose of a mental model enabling the user to predict the operation of a target system. Although, conceptual models are devised as tools for the understanding or teaching of physical systems. Mental models are what people really have in their heads and that guides their use of things.

Apparently, it is important for us to distinguish among several different kinds of models and conceptualizations and the different situations in which they are used. Our conceptualization of a target system should not be confused with the mental model that a user creates of that system. The designer's conceptualization may also differ from the image that the system itself presents to the user. In the ideal world, the system's image will be consistent with the designer's conceptualization, and the user's mental model will also be consistent with both.

If we extrapolate these distinctive models to the building of expert systems it might be like this:

- In this case the target system (t) is the task that must be implemented in an expert system or coherent part thereof.
- The conceptual model of that system C(t). This is the way the task is conceptualized in the expert system. This conceptual model must also fulfill the criteria Norman mentions for his calculator: learnability, functionality, and usability. In case the expert system acquires special features like deductibility, fitting in the overall design of the system, these must be added.
- The user's mental model of that task is M(t).

- The expert conceptualization is C(M(t)), as Norman says, this is often not verbally expressed.

The assignment of the knowledge engineer, who has a mental model M'(t) of the task, is

1. to elicit the expert's conceptualization, and
2. to equalize the mental model of the domain expert, of the future user and his own.

From these data he can construct his conceptualization of the mental model C'(M(t)), which should be on par with the domain expert and the requirements of the expert system.

Summary

Knowledge acquisition and cognitive modeling are necessarily interwoven. Knowledge acquisition shapes cognitive modeling because operational knowledge contains assumptions and directions for its use, that is, an implicit processing model. In return, problem solving models can profoundly shape knowledge acquisition by providing a framework for the articulation and creation of domain expertise.

3.6 CONCLUSIONS

In sections 3.3 and 3.4 I have shown views of human cognition from a theoretical point of view. In section 3.5 I have mentioned problems of human-information processing from a more practical standpoint. The last paragraphs were focused on the cognitive skills of experts. In this section some conclusions are drawn and several proposals of advice for the knowledge engineer are given.

1. We cannot observe the world as it is without interpreting what we experience in our own way. The relation between the world and the knowledge about the world is a contingent and changeable one.

2. Human-information processing deals with the acquisition, representation, and use of knowledge. How this is done is still not known and thus the theories about them are still hypothetical.

3. The physical-symbol hypothesis would lead us to a simple solution of the mismatch problem: human-information processing is an operational, researchable task. There is a single internal representation of information in the mind, where knowledge is made up from primitive elements. Hence, there is no difference between the human and digital-information processes and there is no mismatch. The practice of the knowledge engineer, however, proves this to be untrue.

4. Specially from Norman we learn the shortcomings of man's ability to deal with knowledge. People do not use mental logic, but mental models that are structural analogue to the world, or

that are propositional, or pictorially analogous. The idea is that all this is computable, but also that our knowledge of the world is according to our ability to structuralize it.

5. The way Johnson-Laird presents his mental-models theory is not very convincing. On one hand, he postulates very precisely described but highly hypothetical representations in the case of inference making. On the other hand, he leaves much room for speculation when it comes to the way people make mental representations in general. I doubt very much the form in which Johnson-Laird presents them.

The most useful lesson that can be drawn from Johnson-Laird's work seems that mental models are black boxes in the human mind.
We know that they exist, but we don't know how they work.
People use their own mental model(s) according to their verbal or perceptual abilities and preferences. They also are dependent of education, social context, and background.

6. It is also plausible that people have the ability to use categories as frameworks or schemata for perception and representation. They use concepts to form generalizations, to remember, to learn, to use analogue reasoning, to explain, etc.
However, the situation is similar to the use of mental models, we do not know much about the way people deal with categories and concepts. Like mental models are subject dependent, so are categories and concepts. They are subjective and uncontrollable. One categorization of the same perception is interchangeable for any other, depending on the point of view. Categorization is, while perceiving, abstracting into concepts. It is a pragmatic and economical way to deal with information.

7. Schemata, or organized structures, seem to be used for storing information, but it is far from clear how.

8. When looking at the practice of knowledge engineering, more differences between man and machine are obvious. Man is flexible when dealing with knowledge. He has common sense which he uses most of the time without realizing. He uses a large stock of perceptions as well states of affairs. His knowledge is processed in patterns or schemata, rather than in atoms. And moreover, he has insight and intuition. Until now, it has been impossible to formalize or even structuralize these forms of knowledge.

These findings correspond with the assumptions of the mismatch hypotheses. So far, we have only looked at the human side of the "match". In chapter 4 some knowledge-representation systems will be reviewed. Only then final conclusions about the mismatches can be drawn.

Advice for the knowledge engineer

In the knowledge-acquisition process, the knowledge engineer should be familiar with these aspects and possess a fair amount of psychological insight, in order to be able to ask the following questions.

In the initial state of knowledge acquisition, the mental models of the domain expert should be revealed:

Does he have a verbal, pictorial, or analogue representation of his knowledge?
In what way does he structuralize his knowledge when he communicates about it?
How much common-sense knowledge, intuition, and insight does he use?
Are there particular structures he uses preferably, such as telling cases, drawing flow charts, etc?

In summary, the domain expert's mental models of his expertise must be revealed before the actual knowledge-acquisition process can start. The psychological theories that could be of assistance are still in their infancy, but the foregoing attempts of theory formation have at least pointed out the direction in which expert-system development should continue.

CHAPTER 4 KNOWLEDGE REPRESENTATION

4.1 INTRODUCTION

In the previous chapter human aspects of knowledge have been shown. In order to make both ends of the mismatch problem clear, the topic of this chapter is about knowledge on the machine side.

Since the early years that expert systems were developed, something called "knowledge representation" has been a subject of study. Knowledge representation is the description or image-like reproduction of a state of affairs. In AI programs the way knowledge is represented is crucial for the functioning of the program. The program should be able to draw sound inferences and true and meaningful conclusions. It also must be able to use the knowledge in an intelligent way. These aspects are the main guidelines for a well-formulated knowledge representation.

This chapter's outline will be as follows:

In the next section I shall mention the various methods of knowledge representation. This is necessary, first, to make the reader familiar with the aspects that are involved in knowledge representation. Secondly, it is interesting to see how some theories about knowledge representation are developed with the idea of human-information processing in mind, while others (e.g., logic) regard knowledge representation in terms of powerful inference tools.

This is followed by some of the debates about these methods. As in most young sciences, debates around very basic topics are necessary to pave the way for further research. I have chosen for the logic versus knowledge controversy because it illuminates the mismatch problem from the side of knowledge representation well.

The existence of knowledge representation as an isolated subject might lead us to wrong directions. In 1.4 I have followed Buchanan who says that knowledge acquisition is an ongoing and interactive process in the building of an expert system. From these observations one might conclude that knowledge representation is also a theme that must be regarded in connection with other subjects in AI and the expert-system-building enterprise. In the last section I will elaborate on some new methods in knowledge representation that try to overcome the isolation problem and also look promising regarding the mismatch problems.

4.2 REPRESENTING KNOWLEDGE[31]

"Knowledge" in various forms of representation consists of "data" in various forms of representation. Data are only strings of signs. Data become only knowledge when some agent actively uses the data by giving meaning to the signs. This use can be facilitated according to the structure of the data representation.

The distinction between "data" and "knowledge" is not always sharply made. In the last part of this section I will go further into this matter. But first, I shall mention the various methods of knowledge representation: logic as representation and procedural and declarative representation.

Three kinds of knowledge manipulation can be distinguished in AI:
1. Knowledge acquisition;
2. Retrieving relevant knowledge from the database;
3. Reasoning about it.

These acts are goal directed in order to solve problems.

New data can be added unstructuredly, but it is much better for future retrieval to classify the data and adapt or link them with the existing data. In this way it is easier to retrieve knowledge and to reason about it.

The different kinds of representation reflect the different kinds of reasoning:
Formal reasoning on a syntactic level needs logic as base for representation, but analogous reasoning or inductive reasoning can only be done on a different level in which adequate representations are available. Meta-level knowledge keeps track of the underlying reasoning processes and relevant knowledge, like humans do.

Logic as representation

Logic has been one of the first modes of knowledge representation in AI. Propositional and predicate calculus are being used. Also other kinds of logic, that can express more complicated sentences, are used.

For formal reasoning the representation of facts is made in the form of statements or, better, sentences. Logic is about the form or syntax of statements. The constituents of the sentences can be manipulated so that inferences can be made. All that can be said about the sentences is that they are true or false.

An extension of proposition logic is predicate calculus. With the use of quantifiers one can make inferences about individuals or sets of individuals. Predicates or functions can be used in combination with individuals to represent more complicated expressions, like "city A is between city B and city C".

[31] In this paragraph I rely upon the chapters on knowledge representation in Barr, Cohen, and Feigenbaum (1981 and 1989) and in Rich (1983).

Other kinds of logic are also used for knowledge representation.
E.g., non-monotonic logic that allows to alter statements or to add statements that might contradict others.
In probabilistic logic uncertain statements can be represented.
Fuzzy logic, although not often used in AI programs, can represent properties that are not sharply defined.
Representation of beliefs or sets of beliefs can be given in belief spaces.

Logic programs are very useful as processing tools for solving semi-mathematical problems (e.g., the tower of Hanoi puzzle). It is, however, not a way to represent knowledge when it comes to a heuristic approach to the database. Another disadvantage is that knowledge can seldom be so neatly described that it fits in a logic representation.

Procedural representation

In order to implement heuristic searching, procedural representations in the form of AI programming languages (like LISP) were developed together with extensions in the form of production systems. These systems are made of production rules of the form:
"If <condition> (and <another condition>), then <action>".
The context, or data, or short-term memory buffer, provides the data for the "if" condition. The whole procedure is directed by an interpreter.

Production systems are specially apt to represent knowledge about "how to do". Examples can be found in speech-understanding systems and (medical) diagnosis systems, e.g., MYCIN. Because the rules can be changed independently from the context database the flexibility of the system is better than in network-like systems. In larger systems, like in procedural representation, efficiency and control might decrease.

In sum, logic can hardly be considered as an adequate means of representation. It is only a representation of form. The content or meaning is lost. In the procedural representation is, in addition to the form (If...then...), also represented the content.

A more structured form of meaning representation can be found in networks, frames, and scripts.

Declarative representation

Networks, frames, and scripts are representations that describe objects and their properties. These representations are also known as slot-and-filler structures. By using the human memory as a model, semantic networks give more structure to databases.
In a semantic network it is possible to connect various facts into one scheme of representation. It is easy to add new knowledge to the network by adding objects in nodes and the relations in links. These new facts inherit the properties of the node to which they are attached. For special cases extra information can be added without influencing the other nodes.

The value of such systems can be augmented when the words that are used have a sort of meaning frame. Wilks (1977) uses a set of semantic primitives that defines the words that can be submitted to them. That means that main properties are defined specially for translation. A

system, based on the same principle, but with more applications such as the description of action, can be found in Schank and Abelson's (1977) conceptual-dependency theory.

To grasp complex bodies of knowledge, the idea of representing them in frames or scripts has been developed. According to their proponents, representing events as scripts is similar to the way humans organize their knowledge. Whatever this may be, it is certainly the case that this form of organization of the knowledge has its advantage over other systems: "chunks" of knowledge are easy at hand for recall, inference, and in particular for expectations. Frames are pieces of knowledge, built up with slots that have to be filled with specific knowledge. The frames form together a framework. An example is the Restaurant Frame in which typical restaurant situations are represented. The slots are filled with possible objects and procedures and usually with a default value. Whereas frames consist of declarative knowledge, the slots might be filled with procedures (if...then...).

Frames can be compared with semantic nets, where semantic nets operate on a semantic level, frames can be regarded as being on a pragmatic level. This means that frames are able to deal with the contextual use of language. Frameworks that represent story-like events are called scripts. Scripts represent common sense knowledge about certain events. They can also understand stories, written in natural language, provided that the script and the story are about the same topic.

Analogical representation

Direct or analogical forms of representation, like maps, diagrams, and models, are schemes in which properties such as relations are given as structures. Such representations were already used in Gelernter's Geometry Theorem Prover and other problem solvers with graphic or maplike data. Analogical reasoning is often used in learning programs.
Presently the above mentioned forms are still the basic methods in knowledge representation. Over the years many additions and refinements have been added. They vary according various aspects of knowledge. Some might emphasis the recognition of natural language, others claim better access to the system. Some systems favor logical consistency over heuristic searching.
There are also systems with various levels of knowledge to keep control over the process.

In section 4.3 I will go deeper into the discussions about the various methods of knowledge representation.

The knowledge level

The practice of knowledge representation with the methods mentioned above, looks very much straightforward. However, if we focus on the philosophical issues, we see the same problem occurring with "knowledge" as we did with "understanding" in chapter 2. Therefore this section will be concluded by some remarks on the "knowledge level", an interesting view on the data/knowledge distinction, formulated by Newell in 1982.

To put the various knowledge representation methods in the right view for the mismatch problem, it is necessary to say something more about data and knowledge. Data can be described as passive and uninterpreted "chunks" of symbols. Data can be distinguished from information

which, though also passive, is interpreted, like information we find in books. Knowledge is an extra dimension to information. It is active in the sense that it can transform one or more bits of information into new information, depending on the goal. Knowledge is data plus their processing. In all knowledge-representation systems data and information are represented on a symbol level. Knowledge, however, in this interpretation does not exist on this level. For that reason Newell (1982) introduced the knowledge level as the principle of rationality, that is in terms of a hypothesis, because its existence is so far only a point of view:

"There exists a distinct computer-systems level, lying immediately above the symbol level, which is characterized by knowledge as the medium and the principle of rationality as the law of behavior." (p.99)

New in this approach is the appeal to "rationality". Quoting Newell's explanations:

" - Knowledge is intimately linked with rationality. Systems of which rationality can be posited can be said to have knowledge. It is unclear in what sense other systems can be said to have knowledge.
- Knowledge is a competence-like notion, being a potential for generating action.
- The knowledge level is an approximation. Nothing guarantees how much of a system's behavior can be viewed as occurring at the knowledge level. Although extremely useful, the approximation is quite imperfect, not just in degree but in scope.
- Representations exist at the symbol level, being systems (data structures and processes) that realize a body of knowledge at the knowledge level.
- Knowledge serves as the specification of what a symbol structure should be able to do.
- Logics are simply one class of representations among many, though uniquely fitted to the analysis of knowledge and representation." (p.100)

Apparently, the knowledge level is conceived as an agent with bodies of knowledge and is goal oriented. Therefore it can be regarded as an intentional level like Dennett and Searle[32] do. Actually, they mean something slightly differently, but it might help in understanding what Newell means by his knowledge level (see Newell 1993).
First, he does not hesitate to focus on "intelligence":
A system can only exhibit intelligent behavior if it is described in terms of a symbol and a knowledge level. From this follows the principle of rationality:

" If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action." (p.102)

So his definition of knowledge is:

"Whatever can be ascribed to an agent, such that its behavior can be computed according to the principle of rationality." (p.105)

Finally, Newell defines representation as knowledge plus access to that knowledge. A representation is the structure at the symbol level that realizes knowledge, i.e., it is the reduction of knowledge to the next lower computer-system level.

[32] see Daniel C. Dennett, The intentional stance (MIT Press (1987) 1989) and John R. Searle, The rediscovery of the mind (MIT Press 1992).

In Newell's definition knowledge is still an abstract concept, but Clancey (1989 and 1991) tries to place the knowledge level in a better perspective. He states that only by observing a knowing agent and the description of that knowledge, knowledge turns into a state that can commonly be observed. In other words, knowledge-level descriptions are relative to the observer's point of view:
Interpretation is done by an outside agent. Therefore, knowledge representation is not the knowledge of the agent, and not of the observer. It is a representation of the observer's knowledge of the agent's knowledge. Or in terms of knowledge engineering: representation is the knowledge engineer's knowledge of the domain expert's knowledge. I return to this approach in chapter 5. Here it is sufficient to say that Clancey's investigations on the base of Newell's distinctions result in guidelines which expert-system builders should follow before they proceed to the stage of implementation.

In this section an overview of the possibilities in knowledge representation was given. In the next section I go deeper into the controversies that have occurred between the various ways of knowledge representation.

4.3 THE DEBATES

According to the *Handbook of Artificial Intelligence*[33] *"...there is no theory of knowledge representation."* That is, questions why some representations are better for certain tasks while others are not, are still unsettled. In the last 13 years after this statement was made, it is still not clear why some representations are better than others for certain tasks. Although many publications about topics on theoretical aspects of knowledge representation have been appeared.

In this section I give a sequel to the previously mentioned methods on knowledge representation and the debates that were caused by them. The choice of these debates is also motivated by my interest for the mismatch between human and machine knowledge.

The knowledge-representation hypothesis

Smith (1982) formulates the following hypothesis:
"Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behaviour that manifest that knowledge."

Knowledge of any sort is always theory relative. The representation hypothesis implies that these theories must be explicit. One should consider the fact that computers do not have our implicit (tacit) understanding. As stated in the previous section, it is more and more stressed that the knowledge is at the observer's side. Moreover, knowledge is created by the observer and is itself open to interpretation. This stress of construction of knowledge leads to the constructive view of knowledge.

In the light of these remarks I want to present one of the hottest debates in AI: logic versus knowledge representation.

Logic versus knowledge

A common description of logic is the study of truth preserving inference. Defining a logic is given by a precise specification of the meanings of the well-formed expressions of the logical language. This is done in the semantics of the logic. The meaning of a sentence is the truth value of the sentence, that is, whether the sentence is true or false. Thus, given a possible state of affairs in the world, the semantics determines whether a sentence is true or not.

Logicians, originally, wanted a language such as Frege's within which they could express a set of basic mathematical truths or axioms, from which all the rest could be generated by applying a finite set of precisely prescribed rules of proof that could be shown to be truth-preserving. The

[33] Barr, Cohen, and Feigenbaum 1981, Vol 1, p. 147

languages of the various types of logic were not meant for general use. Their developers did not claim these were universal symbolisms for unrestricted application - that is that everything thinkable could be adequately expressed in them.

Thus, as a representational formalism, a logical language is just a tool. The effectiveness of this tool in carrying out a certain task depends on how it is used. This role of logic as an analytic tool became widely recognized. But many workers in AI rejected the view that logic should also be used as a knowledge-representation language. Applying the definitions of Newell (1982), an intelligent system's behavior is governed by a principle of rationality according to which it will always select those actions that will lead to the attainment of one or more of its goals.
Thus, Newell says that logic is not even the appropriate formalism at the symbol level. He concludes this because of the efficiency problems that led to the failure of the early experiments with the use of logic for knowledge representation.

It is true that logic can give a greater expressive power because there is a large number of logics, each of which allows the formulation of a special type of knowledge. But not all kinds of knowledge can easily be implemented in a certain kind of logic. Notably the predicate calculus, the tool of the mathematical logicians, is basically a context-free formalism; in other words, the meaning of an expression is independent of the context in which it is uttered.
Many of our everyday inferences are also not deductive. In fact, many different kinds of reasoning can be named: inductive, model-based, probabilistic, approximate truth, abduction, which is inference to the best explanation, analogical, default, reducing the number of independent phenomena, counterfactual, etc.[34]
Predicate calculus also cannot make distinctions concerning the closeness of an answer to the truth. Therefore the "notational efficacy" of predicate logic is questioned as will be shown in the following paragraph.

Notational efficacy

Way (1991) discusses in more detail the argument of notational efficacy. She says that the notational efficacy of a representation language concerns not only the actual shape and structure of the representation with respect to its computational efficiency, but also how easy the language is for people to understand and work with (conceptual efficiency). According to Way two aspects of knowledge representation must be considered:

1. Expressive adequacy has to do with the expressive power of the representation - that is, what it can say. Two components of expressive adequacy are the distinctions a representation can make and the distinctions it can leave unspecified to express partial knowledge.

2. Notational efficacy concerns the actual shape and structure of the representation as well as the impact this structure has on the operations of a system. Notational efficacy, in turn, breaks down into such components as computational efficiency, conceptual clarity, conciseness of representation, and ease of modification.

Way endorses the claim that logic does not capture human-reasoning processes. The assumption made by the logicians that a significant amount of thought is deductive is not true.

[34] See McDermott. A critique of pure reason. Computational intelligence (3) 1987

In general, many researchers argue that the structure of semantic networks provides more information and facilitates inferences better than a corresponding representation in first-order predicate logic. Semantic networks explicitly organize information according to a generalization scheme, something which is not clear from a list of implications. Or, as Zadeh (1984) formulated it: logic is of limited or no relevance to AI. First order logic is too limited to be able to deal effectively with the complexity of human cognitive processes. They use ad hoc techniques and heuristic procedures.

Logic versus language

Another difficulty arises in translating natural language into logic. Interesting is that there have never been any claims, at least from logicians, that everything sayable in a natural language could be expressed in predicate calculus. Arguments to the contrary were always countered by "true" logicians, such as Bar-Hillel (1970) who has repeatedly expressed his amazement about the fact that during 2300 years logic has submitted so little to a better formal "grip" on natural language.

Also Israel and Brachman (1986) mention this problem:
Programs for formulating formal semantic account for natural languages use formal languages that are different from that of first-order logic, e.g., Montague or Barwise and Perry's situation semantics. They do not know of any attempt that a fragment of natural language has been semantically analyzed by way of a systematic translation into a standard first-order language.

Conclusion

I conclude recalling Buchanan and Smith' (1988) definition of an expert system:

"An expert system is a computer program that a) reasons with domain-specific knowledge that is symbolic as well as mathematical; b) uses domain-specific methods that are heuristic (plausible) as well as algorithmic (certain); c) performs as well as specialists in its problem area; d) makes understandable both what it knows and the reasons for its answers; and e) retains flexibility."

Even this definition makes it plausible that logic is at most a subsystem of an expert system.
What these sections should have made clear is that the variety of possible other subsystems is still large. This conclusion can also be elucidated with the help of the trade-off between the expressiveness and the tractability of a knowledge-representation system, as explained by Levesque and Brachman (1985).
Neither expressiveness nor tractability by itself can determine the value of a representation language. There is no single best language.

4.4 FRAMES

One of the conclusions of chapter 3 is that humans do not use any kind of formal logic in order to reason. Rather, they use mental models and that sort of pattern-like schemata.

In the previous section I have given enough reasons from within AI for choosing a method of knowledge representation that is more than just logic. From these points of view it seems reasonable to focus our attention to knowledge-representation methods that show more - what I would call - semantic power. Therefore, I shall discuss frames in more detail. They were earlier described as more structured forms of semantic representation. Specially the development of the frame-based language KRL has initiated the success of frames as knowledge representation for knowledge-based systems. In this section frames in general, KRL, and its successors will be discussed.

Normally "frames", "schemes", "schemata", and sometimes "semantic networks" are used without distinction. Sometimes, authors do distinguish them to stress technical differences.

Quillian

Quillian (1968) introduced, as one of the first, semantic networks. His basic idea was to construct a formal representation that could simulate human-semantic memory. In his account of semantic schemes we find nodes or units that represent objects and links that represent information about the nodes. In his PhD dissertation Quillian applied the semantic-network ideas in the field of natural-language understanding: "What sort of representational format can permit the meanings of words to be stored, so that a human-like use of these meanings is possible?"
Unfortunately Quillian's semantic-memory nets, as he had called them, never served the purpose that was intended.

Minsky

Minsky's contribution (1975) is an early theoretical account of frames. He defines a frame as a data-structure for representing a stereotyped situation. A frame is like a network of nodes and relations. The top levels of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have terminals, or slots, that must be filled by specific instances or data. Collections of related frames are linked together into frame-systems. Different frames of a system share the same terminals; this is the critical point that makes it possible to coordinate information gathered from different viewpoints. The frame-systems are linked by an information-retrieval network. Thus, when one frame does not fit the situation, another frame that does fit may be found. Minsky argued that an important function of a frame is to represent a stereotype. The stereotype is an intuitively plausible model of the process by which people fill in information about a situation that is not explicitly mentioned.

In an appendix[35], Minsky states that the development of a well functioning knowledge base is a major intellectual research problem.
He mentions several points of concern (which are partly also reasons why not to use logic):

1. Formalizing the required knowledge. Besides formalizing data also the system must know meta-knowledge facts which are usually common sense, such as cause-and-effect, time, purpose, locality, and types of knowledge.

[35] This appendix is sometimes omitted in later editions of this article.

2. Relevancy and monotonicity. The system must be able to operate with relevant data. It must also be able to select these data in the first place. (This is what is later discussed as the "frame problem" or the "qualification problem".)
3. Procedure-controlling knowledge. A simple deduction cannot deal with a non-transitive concept like nearness: (A near B) & (B near C) does not imply (A near C). If we want to find out about the distance between A and C something like quantification should be applied. The point is that a fixed system can prevent the drawing of simple conclusions.
4. Consistency and completeness. These properties are not necessary for the goals of an intelligent system.

To summarize:

1. Logical reasoning is not flexible enough to serve as a basis for thinking.
2. Minsky doubts the feasibility of representing ordinary knowledge effectively in the form of many small, independently true statements.
3. We need direct ways for linking fragments of knowledge to advice about how they are to be used.
4. Instead of declarative statements we need procedural and structural descriptions.
5. Frames, chunks of knowledge, can represent knowledge better than logic can.

Minsky's theory of frames is based on his view of human thinking.
Thinking begins first with suggestive but defective plans and images that are slowly (if ever) refined and replaced by better ones. When one encounters a new situation (or makes a change in one's view of a problem), one selects from memory a structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary. Minsky does not distinguish between human and machine thinking because neither of them have a soundly explaining theory.

Quillian's and Minsky's approaches, among others, are based on human-information processing. Some of the networks have been proposed as models of human memory and meaning representation, while others are used as components of language understanding and reasoning systems.

The epistemological level of frames

Implementing frames as working systems has encountered difficulties. Brachman (1977) shows the shortcomings of these nets in general. He proposes a better epistemological foundation. The notation of nets does not give a correct reflection of the meanings of the nodes. Therefore, he has tried to add more information and structure explicitly for semantic nets. This work can also be seen as the predecessor of KL-ONE (see p. 60). Moreover, Brachman distinguishes five levels of semantic networks:

1. The implementational level; the level of atoms or mere data, without semantic content. Questions like "What data structures are best suited for representation" are treated on this level.
2. The logical level; the propositions in logical representations, with as extra feature the network topology. Thus, nodes represent predicates and propositions, links represent the logical relationships between these nodes.

3. The epistemological level; this level was specially created by Brachman to describe concept types and knowledge structures.
4. The conceptual level of semantic or conceptual relations. The relation with natural language is close. Nodes represent word-concepts like object, action, or event types.
5. The linguistic level, which is natural language itself.

The formal structure of conceptual units and their relationships forms an epistemology. This epistemological level lies between the logical and conceptual level, that accounts for the internal structures of concepts, their relations and inheritance.
The epistemological level in semantic networks can be described as the set of concept types, inheritance, and structuring relations. This epistemological level should not be confused with Newell's knowledge level which is abstract and relative to the observer.
The epistemological level is independent of the content of the concepts. Concepts represent objects, specially their intensions.

The declarative - procedural controversy

In the same year as Minsky (1975) wrote his theoretical account of frames, Winograd and Bobrow attempted to develop a frame-like representation language, KRL (see Bobrow and Winograd 1977). Before KRL was developed another debate was going on in the AI community. It was about the declarative - procedural controversy. Terry Winograd has straighten out this debate in his 1975 article.

At that time, one could divide the AI community into people who claimed that our knowledge is mainly "knowing that" and others who hold that our knowledge is mainly "knowing how". The declaratives state that we think in data-like structures and the procedurals believe that we think in procedures.

As Hayes (1979) will argue later, Winograd acknowledges here already that in programming languages, like LISP, procedures are taken as data. Or, the other way around, data can be regarded as procedures. So, actually there is no distinction between procedural and declarative statements.
Winograd aims here at the mismatch problem, because he asks the question what we would gain with either of these viewpoints.
"What kind of representation do people use?", and "What kind of representation is best for machine intelligence?" Or better: "What kind do both use?"

The different benefits of declarative and procedural representations are basically that factual knowledge can easily represented in declarative statements and procedures in procedural representations respectively. Thus, both representations, when implemented, work satisfactory in their way. Yet, declarative statements are flexible and economical in use, whereas procedures offer the advantage of interaction between chunks of knowledge and thus more reasoning power.
So Winograd's question was how representations can take advantage of both the abilities of decomposing knowledge in "what" and in "how" and also use the reasoning power. His answer was that this would be possible in frames as representation. From that moment a steady process of constructions of knowledge-representation languages started. Therefore I go more into the details of this process.

KRL

First of all, Bobrow and Winograd (1977) developed KRL, a representation language using frames. It is one of the first systems that mixed procedural and declarative knowledge.
It was originally built as part of a natural-language-understanding system. Its significance lies in the basic intuitions about how knowledge should be organized (p.5):

1. Knowledge should be organized around conceptual entities with associated descriptions and procedures.
2. A description must be able to represent partial knowledge about an entity and accommodate multiple descriptors which can describe the associated entity from different viewpoints.
3. A method of description is a comparison with a known entity, with further specification of the described instance with respect to the prototype.
4. Reasoning is dominated by a process of recognition in which new objects and events are compared to stored sets of expected prototypes, and in which specialized reasoning strategies are keyed to these prototypes.
5. Intelligent programs will require multiple active processes with explicit user-provided scheduling and resource-allocation heuristics.
6. Information should be clustered to reflect use in processes whose results are affected by resource limitation and differences in information accessibility.
7. A knowledge-representation language must provide a flexible set of underlying tools, rather than embody specific commitments about either processing strategies or the representation of specific areas of knowledge.

Thus, declarative knowledge is organized around conceptual entities with associated descriptions. (The authors rely on Rosch' theory of concepts.) Conceptual entities are objects, relations, and events. The system has operations that can augment a description with new knowledge, matching two descriptions, and seeking referents for entities with a specified description.

An - intensional - description is a set of independent descriptors that describe the object from all sorts of viewpoints. It is necessary to sum up all possible descriptors because they will be used in various reasoning processes. The more descriptors the better the system functions. One or several descriptors can serve as prototypical and thus as a default descriptor. The idea behind using prototypes is that people also process their knowledge in prototypes rather than in abstract concepts. Descriptions are clustered into units. A unit or category functions like a concept. The procedure that can be related to a category depends on the category type.

KRL can also represent knowledge about knowledge in the form of meta-descriptors that tell something about a unit of descriptions. Reasoning within these frames is possible by searching for matching descriptors. The descriptors are written in a LISP-like format. Procedures are specified in LISP. This makes KRL a procedural-like language, although the representation is merely declarative. Thus, KRL serves two goals: a programming language (as substitute for LISP) and as a knowledge representation.

Natural language and knowledge representation

In the early '70s much attention was paid to natural-language processing (such as SHRDLU, Winograd 1972). Winograd and Bobrow's KRL was also developed from case grammar. The success semantic networks had in natural-language understanding made these structures so important. But frames became even more important after the conceptual-dependency theory was developed by Schank (1980). This theory enabled the development of an inference engine for Schank's memory structures which was able to handle a much larger and wider range of language than any of the earlier systems. Without going into details, Schank's results were seen as a major step forward in the language-understanding tradition.
AI research became divided into two quite different trends:
Many researchers in language comprehension adopted some form of frame representation, while problem-solving people tended to use predicate calculus. Problem solvers are primarily concerned with deep inferences in narrow domains, while language comprehenders are more concerned with shallow inference in broader areas.
But already in 1981 Charniak suggests a compromise position in which both frames and predicate calculus are used. The advantages of frames that made them so popular for language-comprehension researchers are retained:

1. Frames give a natural way to partition knowledge.
2. A reduction in depth of inference needed during language comprehension.
3. Frames offer a way to make the representation of inferential knowledge more natural.

Critique

In the context of the two different trends in the AI tradition, Hayes' (1979) attack on frames in particular should be mentioned here. His critical analysis of frames boils down to his statement that frames are only another syntax of first-order predicate logic. According to Hayes, all descriptions of concepts can be placed in lists or other kinds of representations and still be part of predicate calculus.
He also mentions problems like different meanings of words according to their frame and the possibility of using defaults. Hayes offers suggestions for the possible solutions, which he calls as "nothing more than syntactic sugar".

Hayes' remarks might be right. However, I think that he misses the point of the rationale behind the frame representation. Frames are developed for representing knowledge, because they resemble - probably - structures in human-knowledge processing. By imitating human-knowledge structures knowledge representation will be much easier and translucent. Moreover, it proved to be possible to develop better techniques for handling frames. I shall mention KL-ONE and KRYPTON.

KL-ONE

Brachman and Schmolze (1985) develop KL-ONE as a result of ongoing research in finding better techniques. They took the advantage of the criticisms on frames. In KL-ONE it is possible to distinguish between descriptions of an object and assertions of their existence. Basic in KL-ONE are the elements or concepts, the structured conceptual objects. Concepts are defined as formal objects used to represent objects, attributes, and relationships. Concepts are descriptions

of objects, and are thus intensional entities. They are distinguished in generic and individual objects. Individual objects are denoted by individual concepts, which are individuations of the appropriate generic concept. The internal structure of an object is defined in so-called roles for the subpieces of the concept, and in structural descriptions for the internal relations.

Expressive adequacy is an important driving force in KL-ONE research, emphasizing the semantics of the representation and its adequacy to make the kinds of subtle distinctions that people make when conceptualizing complex ideas.

KRYPTON

KL-Two, derived from KL-ONE, is also a hybrid system and uses propositional and quantificational reasoning. KRYPTON is their successor. In successful commercial available systems like ART, KEE, and KnowledgeCraft[36] these various representation languages are implemented. They take advantage of all aspects of knowledge representation.

The reasons for their success are made clear by Reichelt (1991):

1. They offer several knowledge-representation languages
2. They offer naturalness of expression
3. They combine frames with logic-based representation which gives the system extra reasoning power.

In addition, KRYPTON has the advantage of being based on a clear epistemology.

One of the features of KRYPTON is the ability to represent data that are not easy to store. KRYPTON offers abstract data structures, with constructors that construct other data structures and selectors which allow to select specified parts thereof. So this language is flexible and can be adjusted for specific data.

Another item is the distinction between assertion and definition. Descriptions can contain information that can be true of a concept like facts (the ABox). They can also give information that serves as a definition or as terminological data (the TBox).
In short, KRYPTON represents the terminology or data information in frame structures. The assertive knowledge is derived from these frames and is put in first-order-predicate calculus. Reasoning can be done using these logical statements.

Conceptual structures

Finally, a representation method should be mentioned that uses schemata and logic in a transparent fashion: conceptual structures. This semantic network-like representation language is one of the few that is not based on KL-ONE. It was Sowa (1984) who presented a theory of conceptual graphs as a method of representing mental models. His claim is:

"Conceptual graphs form a knowledge-representation language based on linguistics, psychology, and philosophy. In the graphs, concept nodes represent entities, attributes, states, and events, and relation nodes show how the concepts are interconnected." (p.69)

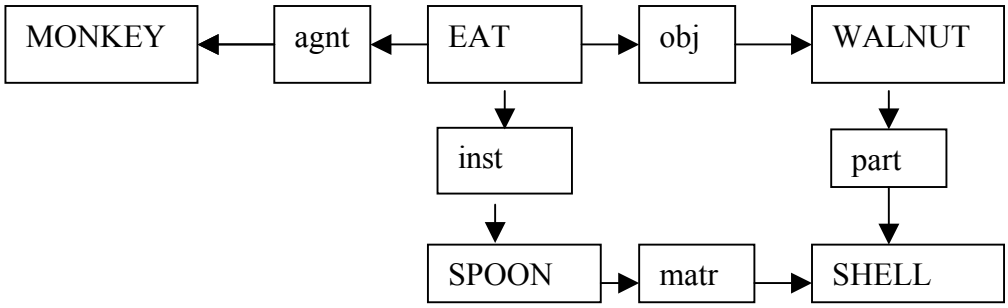
[36] Presently, ART and KEE have been succeeded by new products, based on similar principles. KnowledgeCraft has been discontinued, due to lack of support.

Although the notation is the most visible aspect of conceptual graphs, the complete system is more than a notation: it includes standard operations for logic and computation, conventions for mapping language to and from the graphs, and a semantic theory that relates graphs to real-world situations. As a representation scheme it draws on and integrates ideas from previous work. The result is a more flexible, more extensive and more precisely defined knowledge system than any of its predecessors. The notation gives the full representational power of first-order logic and the mapping onto logic is precisely defined. The notation can also cope with higher order and modal statements.

It was Peirce who invented the existential graph as a system of notation for symbolic logic. In various forms these graphs have been used to represent semantic knowledge in AI. Sowa calls them conceptual graphs and uses them to replace symbolic logic by a more natural notation. They are considered the semantic basis of natural language, representing (mental) models of the world or possible worlds. Sowa's program also tries to formalize the scruffy (see 1.2) kind of AI approach, including a representation for common-sense knowledge, with the use of conceptual graphs.

Sowa's advantage over other representation techniques is that he has ordered the frames and semantic nets in such a way that various sorts of logic can be applied to them. Thus semantic content and syntactic form are presented in one structure.
Another memorable fact is that Sowa gives throughout his book theoretical backing for his findings. So far, only the logical representation methods were based on proofs.

****



********** fig 4.1 A conceptual graph of "A monkey eating a walnut with a spoon made out of the walnut's shell."[37] **********

An important issue is the development of concepts. Sowa gives guidelines for a conceptual analysis:
- Consider every way in which a concept may be used.
- Look for hierarchies; divide instances into types and subtypes.

[37] Sowa (1984) p.78

- Look for links with canonical graphs.
- Check dictionary definitions of the concept and try to complete it.
In addition, one can think of tests to prove the analysis. E.g., by imagining how to teach the concept; how would statements of the concept be tested for their truth or falsehood? At the end a thorough testing for completeness is needed.

The elegancy of conceptual graphs is, besides the representational completeness and the reasoning power, the way instantiations can be made in order to answer questions.

Only in the last few years also implementations of this conceptual-structure representation in working expert systems have been made. Among the adherents of Sowa are Regoczei and Plantinga (1987) who propose the use of conceptual graphs as approximation of natural-language processing. They give the following reasons:
Conceptual graphs stay close to the structure of natural language used by both informant and analyst. And conceptual graphs are a clear notation in which to build models of mental models for public examination. (Like KL-ONE graphs.) Another advantage of Sowa's conceptual-graph notation is that the graphs (in their linear form) are directly machine representable.

Moreover, the authors elaborate on the thought of building an automated-knowledge-acquisition tool using the conceptual-graph notation.
Their hypotheses show a similarity to the psychological hypotheses with which are dealt in chapter 3:

1. Cognitive agents, including computers, understand the world by forming mental models.
2. Mental models have a structure.
3. The structures of mental models can be modeled with conceptual graphs.
4. The operations on mental models can be modeled using operations on conceptual graphs.

They define knowledge as the semantic content of mental models.

According to Regoczei and Plantinga, words "create" knowledge by causing a cogniting agent to form new mental models or altering existing mental models. During the knowledge-acquisition process, the mental models of both expert and knowledge engineer are harmonized; the mental models are "brought together" using language as a technology of harmonization. In this point of view, knowledge acquisition can be reduced to natural-language understanding.

Conclusions

The above mentioned frame-based knowledge-representation methods are just a selection. However, I have tried to make a selection of systems that are promising and, moreover, have set standards[38].

They lead to the following conclusions:

[38] Among others that are not reviewed is SOAR, (see J.E.Laird, A. Newell, and P.S.Rosenbloom. SOAR: An architecture for general intelligence. Artificial Intelligence 33 (1987)) and CENTAUR (see Janice Aikins. Prototypical knowledge for expert systems. Artificial Intelligence 20 (1983)).

1. Frames come closer to the way people deal with knowledge. People have certain concepts of things and events. Over a period of time they add new knowledge to these concepts and alter old knowledge.

2. Frames can be given a rather exact characterization as structures in LISP. In addition, schemata are theoretical psychological entities postulated to explain a variety of observed phenomena of human cognition. So, we cannot expect an exact definition of schemata. A schema is a large, complex unit of knowledge expressing what is typical of a group of instances. In contrast to frames in which both particular objects and general concepts can be represented, schemata involve some sort of abstraction and generalization.

3. When we consider semantic nets as model for human memory and recall the fact that Minsky's theory of frames was built upon the idea how humans think, then we may conclude that frames are - in a way - one of the earlier solutions to the mismatch problem.

CHAPTER 5 SOLUTIONS AND CONCLUSIONS

5.1 BRIDGING THE GAP

The focus of this chapter is aimed at bridging the gap between human and machine knowledge.

In chapter 3 we have seen which aspects are considered significant for human-information processing, such as mental models. In chapter 4 some prominent knowledge-representation methods for AI have been mentioned.
In the last 5 to 7 years a combined research has been accomplished. Some of this work will be reviewed in this chapter.

Given on one side our opaque and vague knowledge about human-information processing and the formal knowledge-representation methods on the other side, all our attention should be drawn to good knowledge-acquisition methods. Knowledge acquisition can be regarded as the intermediate between expert knowledge and the knowledge-based system.

The knowledge engineer has the difficult task to integrate human knowledge into a formal system. Therefore, one of the main tasks of a knowledge engineer is to model the knowledge from the domain expert to the knowledge base. (The other main task is solving problems that arise in the quality of expertise and the communication problems such as resistance, misunderstanding, and access and time constraints with the experts.)

Expert thinking analysis

Regoczei and Hirst (1992) think that it is likely that psychologists will play an active role in the knowledge-acquisition process. Human beings perceive the world around them in a particular and also subjective way. From this perception and the knowledge they already possess, they shape and reshape their knowledge. Here knowledge is built subjectively from parts and pieces of sources unknown to an outsider. The end product of this procedure is the only given fact to the knowledge engineer.
The system on the other hand may store only the end product - in a formalized way - without the previous procedure how it was formed. So far, knowledge is considered to be objective, and the knowledge engineer is expected to be an objective observer.

Knowledge acquisition is not discovery, but more something like creation. The knowledge engineer does not acquire some ready-made knowledge that is available in the head of the domain expert. This means that psychological techniques should be used in knowledge engineering. Fortunately, in the last years a "cognitive shift" has been shown in knowledge engineering.
Besides the classic methods such as interview techniques, protocol analysis, more attention has been drawn to the way of how experts think and reason.

For example, Cooke (1992) proposes that methods from cognitive psychology should be used to retrieve expert knowledge, that is, both the content and representation of expert knowledge. She thinks that in addition to other methods a descriptive approach to knowledge engineering is

necessary. There are many advantages to modeling human expertise. Also, it may be necessary to simulate human cognition in an expert system.

In short, research in cognitive psychology on human expertise points out:

1. The strengths and limitations of expert performance that are relevant to the development of artificial expertise.
2. Expert-novice differences in knowledge representation and search strategies that have implications for the architecture of expert systems.
3. Issues relevant to the expert-system-user interface, such as explanation and automated-knowledge acquisition.
4. Methods applicable to knowledge elicitation. Increased communication between the areas of knowledge engineering and cognitive psychology would not only advance expert-system technology but would also provide a test bed for cognitive theories of expertise. (p.55)

Knowledge-level analysis

As we have seen, an operational definition of knowledge is Newell's knowledge level (1982). He identifies knowledge with goals, aims, objectives, plans, and purposeful behavior (see 4.2). Clancey (1984 and 1989) uses the knowledge level as base for analyzing problems independently from the implementation language that will be used to build the expert system. In his view, 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 systems and its environment, not the internal, physical processes of an isolated system. This analysis is in contrast to implementational-level analysis.

Knowledge-level descriptions cannot be reduced to mechanisms in individual agents: they are relative to the observer's point of view and characterize the total system of agent plus its environment. Knowledge-level descriptions are attributions made by the knowledge engineer involving his own selective interactions with the domain expert, his own perceptions, and his point of view. 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". 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.

Given a knowledge-level analysis, it is possible to devise a better representation language that allows us to make terms, relations, and search procedure explicit. The essence of a knowledge-level analysis is making distinctions - finding patterns - and relating them to the computational process. Rather than talking about goals, they talk about findings (input) and hypotheses (output). Rather than talking about rules, they talk about subtype and causality.

Knowledge-level analysis is a structured way for identifying terms and relations. Clancey's (1984) guidelines:

1. List all possible solutions the program may output; organize into classes and hierarchies if appropriate. Be clear about what the solutions are: plans, processes, configurations, etc.
2. List classes of data that will be input to the program.

3. Identify relations among the data: generalizations, definitions, and qualitative abstraction.
4. Establish the heuristics that link data to solutions after establishing the network of solutions.
5. Treat the search process separately. It is essential to model the expert's inference structure (terms and relations), but not as important to model the search process he uses.

Mental-model analysis

Littman (1989) proposes that we should be able to build intelligent tools to assist the knowledge engineer in building expert systems, if we understand how humans build the kinds of mental models that are represented in the world as expert systems. His research has resulted in three findings:

1. Software designers build and use multiple mental representations when they design software.
2. Software designers make extensive use of mental simulations during the design process to coordinate their multiple mental representations.
3. Expert-system designers exhibit several recurrent mental activities which appear directed toward the construction of mental models.

Not only the designers of systems, but also experts in general use mental models. This use of mental models of the experts to build an expert system is developed by Schumacher and Czerwinski (1992). They have found enough evidence that experts use mental models when dealing with problem solving. A specific feature of mental models is that they schematize knowledge by organizing the knowledge in a special way. The use of mental models implies a better understanding and explaining of a system.

It is not clear whether an expert has one or more mental models of a system or device. The knowledge engineer should pay attention to discovering if there is a mental model and then which mental model the domain expert is using. A mental model is not something that can be put into words, so the knowledge engineer has to look for clues that reveal the mental model(s). Then the mental model needs to be described. To this purpose Schumacher and Czerwinski develop a theory of mental-model acquisition. They base their ideas on memory theories such as found in psychology. In short, they distinguish 3 stages of mental-model acquisition:

1. The pre-theoretical stage,
2. The experiential stage, and
3. The expert stage.

The pre-theoretic stage is the understanding of a system based on recall from memory of superficial similar instances of previously learned systems. In this stage the mental models are a collection of memorized previous experiences.[39] In the experiential stage understanding of causal relationships becomes clearer. In this stage abstraction begins to emerge.[40] In the expert

[39] The authors use among others the categorization theories of Medin and Schaffer to describe how instances are organized in memory. (D.Medin and M.Schaffer. A context theory of classification learning. Psychological Review 85 (1978))

[40] These assumptions are based on MINERVA, a model of memory: D. Hintzman. "Schema abstraction" in a multiple-trace memory model. Psychological Review 93 (1986)

stage the user is able to make abstractions. He can recognize patterns and structures of the system's behavior.
The more a user has accumulated system structures, the easier he can retrieve and compare new systems. His attention has shifted from superficial features towards structural and relational features. The authors propose to use mental models that people have of systems in the design of systems or in user interfaces thereof.

Conceptual analysis

The methods mentioned above are grouped around psychological aspects of expertise knowledge. Whereas these methods depend on psychological insight of the observer, another approach focuses on the linguistic behavior of the expert. This approach, the conceptual analysis, can be regarded as the offspring of conceptual-structures representation (see 4.4).

Sowa (1992) uses conceptual analysis as an aid to formalization. It provides general techniques for analyzing knowledge on any subject. In the formalization stage, the knowledge engineer must encode the knowledge acquired from the expert in the rules of a representation language. Before he can formalize he needs to apply conceptual analysis: the task of analyzing the concepts expressed in natural language and making their implicit relationships explicit.

Understanding of logic and language is a prerequisite for conceptual analysis. Logic is essential because every knowledge-representation language is based on logical rules; language is essential because the primary means of communication between the expert and the knowledge engineer is natural language. In the end, conceptual analysis determines the general principles that define an expert system.

Meaning analysis

In addition to the above mentioned analyses, I propose another form of (conceptual) analysis by decomposing knowledge using the meaning triangle (as in Ogden and Richards, see 2.2)

Three entities or aspects are distinguished:
1. The object or state of affairs as real world entity; the ontological aspect.
2. The name or symbol for this object; the linguistic aspect.
3. The (conceptual) meaning of the object; the conceptual aspect.

We can conceive knowledge as the interrelation between these three aspects.

These entities or aspects can be extended when we regard the object or state of affairs as the representation of that object.
Expert systems are not dealing with the concrete objects, but with representations thereof. Referring to the name of symbol can cause ambiguity. Referring to an object as a representation of that object simulates the perceptive properties. The meanings and intensions people have about objects are the items about which we are communicating. During the knowledge-acquisition process all the three aspects need to be elicited and analyzed. Basically is that the ontological, linguistic, and conceptual approach might differ from person to person.

The conceptual meaning can be compared to the expert's mental model of an object. It serves often as an operational device in problem solving. The expert's conceptual meaning may not be

the same as the user's. So conceptual meanings should be explicitly described in the framework of a knowledge-representation system and also be traceable in the explanation mode of an expert system. Moreover, the conceptual meanings need to be modular as to keep them flexible and conceivable for alteration. Thus, an object can be represented as a frame with slots for attributes and quantitative data. This can be regarded as an internal scheme of the object. The conceptual meaning of an object, the intensions that describe how one regards that object, can also be represented in a frame as the intensional representation of the object. Internal scheme and intensional representation can be placed in one frame but should be distinguished by using labeled slots.

I have not seen knowledge-representation methods that recognize the distinction between a description of an object and its intensional meaning. The attractiveness of representing all aspects lies in the possibility to represent the expert's mental model of an object, his verbalization, as well the description of his perceptual concept of an object. A knowledge-representation method that uses all these aspects forces the knowledge engineer to analyze the knowledge of the expert in a more profound way than systems that do not make this distinction. It will represent more and better (hidden) knowledge of the expert and it is likely that it will perform better.

Summary

Besides the methods that determine the psychology of (expert) thinking in a general way, four other methods look more promising in order to solve the mismatch problem between man and machine, especially in the knowledge-acquisition process.

1. The knowledge-level analysis offers a structured method for revealing terms and relations used by the experts.
2. The mental-model analysis is developed on the evidence that experts use mental models when solving problems. The outcome of the analysis can be used in the knowledge-acquisition process and also in the user interface of the system.
3. A result of the conceptual-structures representation is the use of concepts as cognitive elements in the expert as well in the system.
4. Meaning analysis considers the expert's knowledge from a language (words) and cognitive (meaning) view at the same time.

5.2 CONCLUSIONS

Throughout this paper I have presented evidence that might affirm the mismatch hypothesis. This can be summarized in the following.

The knowledge mismatch

When humans process knowledge in a yet unknown way, how can we know that this knowledge will match constructed knowledge-representation systems? Ample evidence (see chapter 3) has been shown that humans do not use formal rules like inference rules in logic. Most likely they use scheme-like representations or mental models.

In the last years representation languages have been developed that also use scheme-like representations. The initiative for developing frames was in fact human-information processing. So this common base would lead to approximate human and machine knowledge. However, there is much left open. Regarding justification of human reasoning little can be said. Knowledge-representation systems are still operating with logic as inference machine. On this point the mismatch will probably remain until more flexible kinds of logic are available. The question when and which logic should be used is depended on the situation.

The knowledge mismatch is partly solved by frame representation and conceptual graphs. The methods of reasoning are still on different tracks.

The representation mismatch

Two important aspects are here prevalent: common-sense knowledge and the verbalization of knowledge. Since the system CYC (Lenat and Feigenbaum 1991) is supposed to be a general data-base for common-sense knowledge, future machines can be provided with that kind of knowledge. Yet, questions about the effectiveness and easiness of access will remain (the frame problem).

Natural language, specially the relation between language and thinking can never be formalized without the aid of a human interpreter. Though, with techniques such as conceptual and meaning analysis less misunderstanding will occur.

I have left open the question if machine-learning methods and connectionism might develop into a solution.

The completeness mismatch

This mismatch is specially relevant in automated-knowledge-acquisition systems. The hidden or tacit knowledge of an expert will never be revealed unless it is explicitly asked. But how can the system (or even the human analyst) know for what to ask? Special interviewing techniques are needed. Also during the testing of an expert system in development this hidden knowledge can be revealed. Not only tacit knowledge is a typical human attribute that is hard to reveal, also insight and intuition are forms of knowledge that can hardly be expressed. The completeness mismatch looks like an essential discrepancy between man and machine.

The category mismatch

The category mismatch is another essential discrepancy between man and machine. In chapter 2 I have elaborated on one of the basic differences: the lack of understanding and, moreover, the lack of intentionality in machines. The twain shall never meet.

Solutions

Solutions work only in limited and definable areas of information processing. As is shown in section 5.1, knowledge-representation methods that simulate approximately human-knowledge

representation are promising examples. In addition to representation methods, analysis of human thinking is of crucial importance. This analysis can be distinguished in the analysis of the subjectivity and relativity of human knowledge; and the analysis of human knowledge in terms of the language and the meaning he uses in his dealing with knowledge.
In these techniques man and machine reach out and touch.

****

BIBLIOGRAPHY

Arnheim 1969 (1974)
Rudolf Arnheim. Visual thinking. (U. of California Press 1969, 1974)

Bar-Hillel 1970
Y. Bar-Hillel. Aspects of language. (The Magnes Press. Jerusalem 1970)

Barr, Cohen, and Feigenbaum 1981
Avron Barr and Edward A.Feigenbaum, eds. The Handbook of artificial intelligence. Volumes I and II. Paul R.Cohen and Edward A.Feigenbaum, eds. Volume III. (William Kaufmann 1981)

Barr, Cohen, and Feigenbaum 1989
Avron Barr, Paul R.Cohen, and Edward A.Feigenbaum, eds. The Handbook of artificial intelligence. Volume IV. (Addison-Wesley Publishing Company, Inc. 1989)

Bechtel and Abrahamsen 1990
William Bechtel and Adele A. Abrahamsen. Beyond the exclusively propositional era. Synthese 82 (1990)

Berliner 1980
Hans Berliner. Computer backgammon. Scientific American 242 (1980)

Bobrow and Winograd 1977
Daniel G.Bobrow and Terry Winograd. An overview of KRL, a knowledge representation language. Cognitive Science 1 (1977)

Boden 1977
Margaret A.Boden. Artificial intelligence and natural man. (Basic Books, Inc., New York 1977)

Boden 1989 (1988)
Margaret A.Boden. Computer models of mind. Computational approaches in theoretical psychology. (Cambridge U.P. 1988, 1989)

Brachman 1977
Ronald J.Brachman. What's in a concept: structural foundations for semantic networks. International Journal of Man-Machine Studies 9 (1977)

Brachman 1979
Ronald J.Brachman. On the epistemological status of semantic networks. In N.V.Findler. Associative networks; representation and use of knowledge by computers. (Academic Press NY 1979)

Brachman and Levesque 1985
Ronald J.Brachman and Hector J.Levesque, eds. Readings in knowledge representation. (Morgan Kaufmann Publ., Inc. 1985)

Brachman and Schmolze 1985
Ronald J.Brachman and James G.Schmolze. An overview of the KL-ONE knowledge representation system. Cognitive Science 9 (1985)

Brand and Harnish 1986
Myles Brand and Robert M.Harnish, eds. The representation of knowledge and belief. (University of Arizona Press 1986)

Buchanan et al. 1983
Bruce G.Buchanan, et al. Constructing an expert system. In: Hayes-Roth, Waterman, and Lenat 1983

Buchanan and Smith 1988
Bruce G.Buchanan and Reid G.Smith. Fundamentals of experts systems. Annual Review of Computer Science 3 (1988)

Cassirer 1975 (1955)
Ernst Cassirer. The philosophy of symbolic forms. 3 Vls. (Yale University Press 1955, 1975)

Charniak 1981
Eugene Charniak. A common representation for problem-solving and language-comprehension information. Artificial Intelligence 16 (1981).

Charniak and McDermott 1985
Eugene Charniak and Drew McDermott. Introduction to artificial intelligence. (Addison-Wesley Publ. Comp. Inc. 1985)

Clancey 1984
William J.Clancey, Knowledge acquisition for classification expert systems. Heuristic Programming Project; Working Paper (HPP 84-18) Stanford University, July 1984.

Clancey 1989
William J.Clancey. The knowledge level reinterpreted: modeling how systems interact. Machine Learning 4 (1989)

Clancey 1991
William J.Clancey. The frame of reference problem in the design of intelligent machines. In: Kurt Vanlehn, ed. Architectures for intelligence. (LEA, Hillsdale NJ, 1991)

Clark 1989
Andy Clark. Microcognition: Philosophy, cognitive science, and parallel distributed processing. (The MIT Press 1989)

Compton and Jansen 1990
P.Compton and R. Jansen. A philosophical basis for knowledge acquisition. Knowledge Acquisition 2 (1990)

Cooke 1992
Nancy J.Cooke. Modeling human expertise in expert systems.
In: Hoffman 1992

Dennett 1986
Daniel C.Dennett. The logical geography of computational approaches: a view from the East Pole. In: Brand and Harnish 1986.

Doorman 1989
S.J.Doorman. Methodologische opmerkingen over een gedachten-experiment. Kennis en Methode 13 (1989)

Dreyfus 1979
H.L.Dreyfus. What computers can't do. (New York 1972, rev. ed. 1979)

Edwards 1972 (1967)
Paul Edwards, ed. The Encyclopedia of philosophy. (Collier and Macmillan, Inc. New York 1967, 1972)

Feigenbaum 1977
E.A.Feigenbaum. The art of artificial intelligence. In: Proceedings of the 5th IJCAI. 1977

Feigenbaum and Feldman 1963
E.A.Feigenbaum and J.Feldman. Computers and Thought. (McGraw Hill 1963)

Fetzer 1988
James H.Fetzer, ed. Aspects of Artificial Intelligence. (Kluwer, Dordrecht 1988)

Fodor 1975
Jerry A.Fodor. The language of thought. (The Harvester Press 1975, 1976)

Fodor en Pylyshyn 1988
Jerry A.Fodor and Zenon W.Pylyshyn. Connectionism and cognitive architecture: a critical analysis. Cognition 28 (1988)

Frege 1969 (1892)
Gottlob Frege. Ueber Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik NF 100 (1892). Also in: Funktion, Begriff, Bedeutung. Fünf logische Studien. (Vandenhoeck & Ruprecht, Göttingen 1969)

Gentner and Stevens 1983
D.Gentner and A.L.Stevens, eds. Mental Models. (Erlbaum 1983)

Haugeland 1981
John Haugeland, ed. Mind design. (MIT Press 1981, 1982)

Hayes 1979 (1985)
Patrick J.Hayes. The logic of frames. In: Brachman and Levesque 1985. Orig. 1979

Hayes-Roth, Waterman, and Lenat 1983
Frederick Hayes-Roth, Donald A.Waterman, and Douglas B.Lenat. Building expert systems. (Addison-Wesley Publishing Comp. 1983)

Hoffman 1992
Robert R.Hoffman, ed. The psychology of expertise; cognitive research and empirical AI. (Springer 1992)

Holland et al 1986 (1989)
J.H.Holland, K.G.Holyoak, R.E.Nisbett, and P.R.Thagard. Induction, processes of inference, learning, and discovery. (MIT Press, 1986, 1989)

Israel and Brachman 1986
David J.Israel and Ronald J.Brachman. Some remarks on the semantics of representation language. In: M.L.Brodie, John Mylopoulos and Joachim W.Schmidt, eds. On conceptual modeling. (Springer 1986)

Johnson-Laird 1980
P.N.Johnson-Laird. Mental models in cognitive science.
Cognitive Science 4 (1980)

Johnson-Laird 1983 (1987)
P.N. Johnson-Laird. Mental models. Towards a cognitive science of language, inference, and consciousness. (CUP 1983, 1987)

Johnson-Laird 1989a
P.N.Johnson-Laird. Human experts and expert systems. In: Murray and Richardson 1989

Johnson-Laird 1989b
P.N.Johnson-Laird. Mental models. In: Posner 1989

Johnson-Laird 1991
P.N.Johnson-Laird. Deduction (Erlbaum 1991)

Johnson-Laird and Watson 1977 (1985)
P.N.Johnson-Laird and P.C.Watson, eds. Thinking; Readings in cognitive science. (CUP 1977, 1985)

Kitcher 1990
Patricia Kitcher. Kant's dedicated cognitivist system. In: J-C. Smith, ed. Historical foundations of cognitive science. (Kluwer Academic Publ. Dordrecht 1990)

Lenat and Feigenbaum 1991
D.B.Lenat and E.A.Feigenbaum. On the thresholds of knowledge. Artificial Intelligence 47 (1991)

Levesque and Brachman 1985
Hector J.Levesque and Ronald J.Brachman. A fundamental trade-off in knowledge representation and reasoning. In: Brachman and Levesque 1985

Littman 1989
David Littman. Constructing expert systems as building mental models; or Toward a cognitive ontology for expert systems.
In: K.Morik, ed. Knowledge representation in machine learning (Springer 1989)

Miller 1981
George A.Miller. Trends and debates in cognitive psychology. Cognition 10 (1981)

Minsky 1968
M.Minsky, ed. Semantic information processing. (MIT Press 1968)

Minsky 1975 (1985)
Marvin Minsky. A framework for representing knowledge.
In: Brachman and Levesque 1985. Orig. in P.H.Winston ed. The psychology of computer vision (NY 1975). Also in Haugeland 1981

Murray and Richardson 1989
Linda A.Murray and John T.E. Richardson, eds. Intelligent systems in a human context. (OUP 1989)

Neisser 1976
Ulrich Neisser. Cognition and reality. (W.H.Freeman and Comp. San Francisco 1976)

Newell 1982
Allen Newell. The knowledge level. Artificial Intelligence 18 (1982)

Newell 1993
Allen Newell. Reflections on the knowledge level. Artificial Intelligence 59 (1993)

Newell and Simon 1963
A. Newell and H.A.Simon. GPS, a program that simulates human thought. In: Feigenbaum and Feldman 1963

Newell and Simon 1981 (1975)
A.Newell and H.A.Simon. Computer science as empirical inquiry: symbols and search. In: Haugeland 1981. Also in Communications of the Association for Computing Machinery 19 (1975)

Norman 1980
Donald A.Norman. Twelve issues for cognitive science. Cognitive Science 4 (1980)

Norman 1981
Donald A.Norman. Perspectives on cognitive science. (Ablex Publ. Corp. Norwood, NJ. 1981)

Norman 1983
Donald A.Norman. Some observations on mental models. In: Gentner and Stevens 1983

Normore 1990
Calvin G.Normore. Ockham on mental language. In: J-C. Smith, ed. Historical foundations of cognitive science. (Kluwer Academic Publ. Dordrecht 1990)

Ogden and Richards 1972 (1923)
C.K.Ogden and I.A.Richards. The meaning of meaning. A study of the influence of language upon thought and of the science of symbolism. (Routledge & Kegan Paul Ltd. London 1972)

Osherson & Smith 1990
Daniel N.Osherson and Edward E.Smith. An invitation to cognitive science. 3 Vls. (The MIT Press 1990)

Polanyi 1967
M.Polanyi. The tacit dimension. (Doubleday. New York 1967)

Posner 1989
M.I.Posner, ed. Foundations of cognitive science. (MIT Press 1989)

Pylyshyn 1978
Zenon W.Pylyshyn. Imagery and artificial intelligence. In: Savage 1978

Quillian 1968
M.R.Quillian. Semantic memory. In: Minsky 1968

Rapaport 1988
W.J. Rapaport. The Korean Room. In: Fetzer 1988.

Regoczei and Hirst 1992
Stephen B.Regoczei and Graeme Hirst. Knowledge and knowledge acquisition in the computational context. In: Hoffman 1992

Regoczei and Plantinga 1987
Stephen Regoczei, Edwin P.O.Plantinga. Creating the domain of discourse: ontology and inventory. International Journal of Man-Machine Studies 27 (1987)

Reichgelt 1991
Han Reichgelt. Knowledge representation; an AI perspective. (Ablex, Norwood, NJ 1991)

Rich 1983
Elaine Rich. Artificial intelligence. (McGraw-Hill International Book Company 1983, 1991)

Rips 1986
Lance J.Rips. Mental muddles. In: Brand and Harnish 1986

Rorty 1979
Richard Rorty. Philosophy and the mirror of nature. (Princeton U.P. 1979)

Rosch 1978
Eleanor Rosch. Principles of categorization. In: E.Rosch and B.B.Lloyd, eds., Cognition and categorization. (Hillsdale 1978)

Savage 1978
C.W.Savage, ed. Perception & Cognition. Issues in the foundations of psychology. Minnesota Studies in the Philosophy of Science 9 (1978)

Schank 1980
Roger C.Schank. Language and memory. Cognitive Science 4 (1980)

Schank and Abelson 1977
R.C. Schank and R.P.Abelson. Scripts, plans, goals and understanding (Hillsdale NJ 1977)

Schumacher and Czerwinski 1992
Robert M.Schumacher and Mary P. Czerwinski. Mental models and the acquisition of expert knowledge. In: Hoffman 1992

Searle 1980
John R.Searle. Minds, brains, and programs. The Behavioral and Brain Sciences 3 (1980). Also in: Haugeland 1981

Shortliffe 1976
E.H.Shortliffe. Computer-based medical consultations: MYCIN. (American Elsevier, New York 1976)

Simon 1978
H.A.Simon. On the forms of mental representation. In: Savage 1978

Simon 1989
H.A.Simon. Models of thought II. (Yale Univ. Press, New Haven 1989)

Slatter 1987
P.E.Slatter. Building expert systems: cognitive emulation. (Ellis Horwood 1987)

Sloman 1971
Aaron Sloman. Interactions between philosophy and artificial intelligence: the role of intuition and non-logical reasoning in intelligence. Artificial Intelligence 2 (1971)

Sloman 1983
Aaron Sloman. Under what conditions can a machine attribute meaning to symbols. In: Proceedings of the IJCAI 1983

Sloman 1985
Aaron Sloman. What enables a machine to understand? In: Proceedings of the IJCAI 1985

Smith 1982
Brian C. Smith. Prologue to "Reflection and semantics in a procedural language". PhD diss. (Cambridge MA 1982). Also in Brachman and Levesque 1985

Smith 1989
Edward E.Smith. Concepts and induction. In: Posner 1989

Smith 1990
Edward E.Smith. Categorization. in: Osherson and Smith 1990.

Sowa 1984
J.F.Sowa. Conceptual structures; Information processing in mind and machine. (Addison-Wesley 1984)

Sowa 1992
John F.Sowa. Conceptual analysis as a basis for knowledge acquisition. In: Hoffman 1992

Visser 1985
H.Visser. Kennistheoretische problemen van kunstmatige intelligentie. Tilburg University 1985

Vlaanderen 1990
Marie José Vlaanderen. Automated knowledge acquisition for expert systems; an overview. Report Faculty of Philosophy. Erasmus University Rotterdam 1990

Way 1991
Eileen Cornell Way. Knowledge representation and metaphor.
(Kluwer 1991)

Weizenbaum 1966
Joseph Weizenbaum. ELIZA - A computer program for the study of natural language communication between man and machine. Communications of the ACM 6 (1966)

Weizenbaum 1976
Joseph Weizenbaum. Computer power and human reason. From judgement to calculation. (Freeman and Co 1976)

Wilks 1977
Y.A.Wilks. Methodological questions about artificial intelligence: Approaches to understanding natural language. Journal of Pragmatics 1 (1977)

Winograd 1972
Terry Winograd. Understanding natural language. (Academic Press, Inc. New York 1972)

Winograd 1975 (1985)
Terry Winograd. Frame representations and the declarative/procedural controversy. In: Brachman and Levesque 1985. Orig. 1975

Winograd and Flores 1987
Terry Winograd and Fernando Flores. Understanding computers and cognition: a new foundation for design. (Ablex Publ. Corp. Norwood, NJ 1987, 1986)

Woodward 1992
J.Brian Woodward. Developing K-ONCOCIN: a case study in the cognitive process of knowledge engineers. Knowledge Acquisition 4 (1992)

Zadeh 1984
L.A.Zadeh. Interview in The Communications of the ACM april 1984