15-01-2010

MENTAL MODELS IN HUMAN-COMPUTER INTERACTION

Niels Ole Bernsen, Roskilde University

Abstract: The paper discusses a subset of existing theoretical and experimental results on mental models and their application to problems in human-computer interaction. Mental models research is considered as representing a step towards a more fine-grained study of user models than standard HCI studies of human competence and performance. The logical next step in this direction, it is argued, is to focus on the development of a "procedural semantics" of mental models dealing with the construction and use of mental models in the context of computerized work.

1. Approaches to Computerized Work Design.

Design of computer systems for work comprises the analysis and design of advanced information technology mediating between humans and their work including the design of interfaces. In order to scientifically address the problem of design of human-computer systems for work one has to face, firstly, the complexity of the HCI problem involving work domain and tasks, system, interface, user, and methodology of approach. Secondly, results have to be applied to the design process which, being a computerized work domain by itself, shares all of the complexity just mentioned. Thirdly, theoretical foundations are unstable and fragmentary. Forthly, the technology of systems and interfaces is in continuous development and continues to pose new versions of the HCI problem to science and design alike. The complexity of the HCI problem is part of the explanation why, so far, there are relatively few examples of substantial direct impact from scientific theory on product design. Most approaches have focused on a limited number of aspects of the overall problem complexity to the effect that their results are not directly or substantially applicable to the design context in which some version of the entire complexity is at stake. Furthermore, even when applicable to aspects of a given design context, such results may not be of any real constructive use to the designer and tend not to be easily generalizable to different contexts or to larger-scale contexts.

The HCI problem complexity facing designers and scientists alike may be illustrated as follows:

Complexity of work domain and tasks:

The task domain involves: environment, organisation and people, the work to be done in context, and a number of tasks which can be identified more or less precisely. By definition, persons carry out their work using computerized equipment. There are many different work domains characterized by instantiating very different combinations of the above generic components which need to be studied both individually and in an integrated fashion.

Complexity of system and interface:

There are many different types of interactive computer system such as expert systems, decision support systems, word processors, e-mail systems, public service systems, process control systems, databases, programming systems, tutoring systems, entertainment systems. Each such system has an interface to the user through which the user interacts with the system to do work. The interface has its own complexity as determined by the contents of a specific configuration of components such as keyboard, mouse, screen, icons, menus, and so on. To take a simple example, electronic mail systems tend to have rather different interfaces and a thorough and perspicuous description of any one of these at relevant levels of abstraction as a basis for design discussions or scientific study tends to be quite complex. Furthermore, the "system" may be double in the sense of comprising both a computer system and a second system to be controlled through the computer system, as in the domain of process monitoring and control (production plants, power plants, air traffic control, etc.).

Complexity of user:

It is difficult to access and to model in detail the complex mechanisms of perception, learning, knowledge representation, inference and reasoning, goal structures, motor coordination, linguistic skills, etc., in the user relevant to the HCI problem. Current approaches address this problem at very different levels of detail and along different problem dimensions. Users may be novices or experts with respect to the system at hand and have, in addition, different individual characteristics including differences in education and training and various types of more individual differences (in background knowledge, problem solving abilities, etc.).

Complexity of methodology of approach:

Many different experimental, observational, and conceptual approaches having different grain sizes and different theoretical backgrounds are used in trying to address the HCI problem.

The design context:

The design context constitutes a work domain in itself possessing all of the above complexity. Its core is an iterative process of refinement and optimisation (or tradeoff) using feedback at various points in the development process which involves the stages: system requirements => system specification => system implementation, and which requires the taking into account of numerous additional constraints such as time pressure and economy. Some system requirements, though of obvious relevance to human-computer interaction, are less relevant to the study of the HCI problem as they are studied almost exclusively within a computer science framework. Examples are: system consistency, system completeness, system reliability.

The problem domain of systems design may be summarised in the following picture:

Actual Design Process (s)

(Optimization)

Designer's know-how (s,g):

(Empirical rules and heuristics)

Design Guidelines and Standards (s,g)

(Empirical rules and heuristics)

Design Rationale (s)

(Questions, options, criteria)

Tools for Prediction and Testing (s):	Conceptual Tools (g):
- Prototype user testing results - Psychological experimentation (domain and task simulation results) - Predictive user models - Predictive system models - Scenarios for system use -	- Taxonomy of work domains - Skill-, rule-, knowledge based control - Taxonomy of mental representations - Taxonomy of tasks - Error theory - Taxonomy of systems - Learning theory - Frameworks for user and system modelling - Theory of cooperative work - Theory of design - Empirical and modeling methodologies - . . . .

Note: s denotes specific knowledge, g denotes general knowledge.

This picture is obviously incomplete but it does illustrate some structural points. A typical design process solves the design optimisation task using the designer's know-how, design guidelines and standards, prototyping, and, possibly, some more or less systematic user testing and ditto scenarios for system use. This process is neither arational nor uninfluenced by scientific findings but it is less articulate, explicit, systematic, and predictive than is desirable. What science can do is to support the design process through the provision of more explicit tools for testing, prediction and articulation. The design rationale approach (e.g., MacLean, Bellotti, and Young 1990) can help focusing on salient issues in design; techniques for domain, task, user, interface, and system simulation can help predicting the outcome of different design decisions; and conceptual tools from HCI-related theory can help in identifying, clarifying, and solving design problems.

2. Some Results (A): The SRK Framework and Associated Mental Representations.

To locate the possible role of mental models within this complex problem domain, we may start from a description of the skill-, rule-, and knowledge-based (SRK) framework. This framework is a global model of cognitive control based on analyses of performance in complex work situations (Rasmussen 1983,1990). Analysis has identified three prototypical levels of representation of work context. The level which is active in a given work situation depends on the person's background, training, goals, and task, as well as on the system involved.

Skill-based performance is sensori-motor performance during work. An example is the skilled driving of a car. Typically, skill-based performance gets intentionally "released" through a decision to achieve a certain result and, once released, requires little conscious attention as it unfolds in a familiar context. Performance is smooth, automated, integrated, real-time activity which is synchronized with a changing work context via perceived signs. Such performance is strongly based on feed-forward control in which patterns of actions are composed from subroutines to comply with the demands of the situation. Performance may be accompanied by anticipatory higher-level control in general terms but agents have only limited awareness of the perceived sign-action relationships involved. For obvious reasons, the mechanisms of skill-based performance are difficult to identify through standard psychological methods.

A plausible, though highly general explanation of skill-based performance is that it is based on a flexible and efficient "dynamical world model" (Rasmussen 1990). The term "model" refers to the information structures in the agent which in some way have to represent the agent in familiar environments structured in terms of perceptually accessible properties relevant to standard goals. The person executes performance through skilled sensory-motor control as long as no incoming information violates the dynamic world model. As long as this is the case, the dynamic world model directs attention, controls information gathering, and transforms the initial intention into movement control. The information structures involved may bear a certain resemblance to AI constructs such as frames, scripts, stereotypes, or scenarios, but they are probably also very different from these. They are like, say, the perceptual-motor invariances constructed by an infant familiarizing itself with some local environment. On the other hand, it seems clear that the skill-based level of performance and its related dynamic world models are active in, and basic to, all human performance.

Rule-based performance is performance attentively controlled by a stored rule or procedure. The rule or procedure to apply is selected on the basis of perceived cues in the environment. Performance has the explicit form: if this is the case then do that. This means that the person is able to formulate the rules involved. Rule-based performance may involve explicit goals as in: "If I want the computer (or system) to achieve result R, and it is in state S, then I have to do action A". Rule-based performance is mostly based on past I/O observations. I/O rule-based performance may turn into skill-based performance given enough practice and with an accompanying loss of explicit knowledge of the rules involved. A person possessing only cue-action knowledge (and possibly skill-based knowledge) of some system in the work domain is only to a very limited extent able to perform problem solving on the system by, e.g., inferring predictions of system performance under unfamiliar input conditions or explaining system behaviour. The mentioned, explicitly goal-directed rule-based performance belongs to a larger family of system and task related rule-based behaviour also including, e.g., cue-expectation rules of the forms: "if this happens then expect that to happen", "system behaviour B is a reliable cue that the system is in the internal state S", and cue-action-expectation rules of the form: "if I do this and the system is in such-and-such a state then that will happen".

Knowledge-based performance is, (again) prototypically, problem solving performance controlled by attention to explicit goals and based on explicit models of some domain. Examples are prediction of system behaviour and troubleshooting based on explicit system models. Such performance requires the retrieval or construction of an appropriate model, reasoning within the model, and, possibly, subsequent planning and action based on the conclusions obtained. Knowledge-based performance is resource demanding and is used primarily when goals cannot be achieved through skill-based or rule-based performance alone. Once an appropriate model has been constructed and possibly externalized (mathematically, as a graph or diagram, or as a physical model, possibly on the display), skill-based or rule-based performance may resume dominance in controlling operations on the model. Obviously, knowledge-based performance relies not only on models of the specific system at hand, but also on general background knowledge of the world.

Relationships within the SKR framework: There is dynamic interaction between skill-, rule-, and knowledge-based performance both during the performance of individual action sequences and during learning. In both conditions, mismatches within and between levels constitute a prominent source of human error (Rasmussen 1990). Skills may be learned either directly or on the basis of rule-based performance, or even departing from knowledge-based performance. A classical discussion concerns whether skill-based performance, when somehow "derived" from rule-based control through training, is merely a compilation of fine-tuned rules (Anderson 1983), or whether the transition from rule-based to skill-based performance requires a qualitative shift to some more holistic kind of control involving aggregation of hitherto separate pieces of performance, a declarative-procedural shift in knowledge representation, and a shift in the nature of perceived information from cues to signs (Rasmussen 1990). Some results concerning the dynamic interaction between the three levels of performance in the carrying out of specific tasks have been summarized in Rasmussen's decision-ladder (Rasmussen 1986). The skill-based level and its "dynamical world model" are at one and the same time fundamental to the other levels of performance, close to perception and motor action, and highly inarticulate and introspectively inaccessible. These features are important to bear in mind when discussing mental model theory.

3. Some Results (B): System Knowledge.

A system may work more or less exclusively at the skill- and rule-based levels. In such cases, which are quite frequent, the operator does indeed have system knowledge in the sense of skill-based and rule-based knowledge. But the operator does not necessarily have any elaborate models of the system which, in the limit case, may be considered purely as a black box. The absence of such models significantly limits the kinds of tasks which the operator is able to perform on the system(s): problem-solving tasks such as explanation of system behaviour beyond the I/O level, knowledge-based diagnosis of system faults, ditto planning of intervention in the system's operations, ditto prediction of the results of an intervention, or prediction of system behaviour in new circumstances, cannot be performed with any reliability. The limiting case of the purely rule-based operator only has the shallow rule-based substitutes available for such tasks. However, it is probably inevitable that such operators do form models of the system of some kind or other through inferences based on the information provided by the system interface (which, in addition to task-related information, often contains references to internal system processes and states), observations of system I/O, analogies with other, more well-known systems, general background knowledge, previous theoretical education, and so on. But these models may have little to do with the actual form and functioning of the system and may lead to frustration and error during operation of the system due to negative interferences with the interpretation of I/O properties of the interface. Furthermore, in many systems, such as process control, anomalies arise all the time and the cost of having operators who are able only to apply I/O knowledge to system operation may easily become too high. Given the inevitability of model-formation even in users whose tasks are confined to system I/O, it is a major question in interface design whether the interface (or system image, see below) should in general support the formation of models of the internal functioning of systems and, if so, at what levels. We shall see later that mental system models can also facilitate procedural learning.

Conversely, if tasks other than operating the system at I/O level are to be reliably performed then the operator does need to have knowledge of the system at the knowledge-based level. It is important to realize how complex such system knowledge can be. The implication is that a system model in the sense of an integrated model comprising all aspects of the system down to an arbitrary level of detail is something that is neither needed nor had by any system operator. This raises the question of which kinds of system knowledge and models are relevant and sufficient to the various tasks performed by system operators.

Let us look at some of the dimensions involved in relevant system knowledge. Two such dimensions, i.e., part-whole and means-ends relationships, have been investigated by Rasmussen (1979,1986) in studies of troubleshooting in hardware systems. Such systems can be broken down into a structure of elements such as: whole system, subsystems, sub-subsystems, and so on, down to individual system components. In addition to this part-whole decomposition of the system into elements, each such element can be considered and described at a number of different levels in a means-ends hierarchy. At the lowest level, the system element is described in terms of its physical form. A component is, e.g., described as having a specific location in the system and consisting of a specific configuration of materials which have to satisfy such and such demands as to size, purity, etc. At the second level, a system element is described with respect to the way it operates in order to provide a functional contribution to the working of the overall system. At the third level, the element is described in terms of its generalized functionality. Pumps, for instance, may be implemented in very different materials and operate in very different ways while performing one and the same general function in the overall system. At the fourth level, focus shifts to the abstract functionality of system elements in supporting the general coordination and flow of energy, information, or material through the system and hence to aspects of general system connectivity, spatio-temporality, and causality. Finally, at the fifth level, system elements are being considered and described in terms of their overall purpose. This is often where troubleshooting starts, that is, when it is observed that the system does not function properly or does not respond properly to the operator's input.

During troubleshooting, focus of attention moves back and forth within the part-whole system decomposition as well as up and down within the means-ends hierarchy accompanied by conceptual shifts in system description. The major conceptual shift within the means-ends hierarchy is that between physical system description and functional description of the purposes served by the system and its elements. During this goal-directed problem solving process of moving around in the problem space determined by the part-whole and means-ends hierarchies, many different types of mental representations and reasoning strategies tend to be applied, including spatial reasoning, temporal reasoning, causal reasoning, property-inheritance reasoning, reasoning by analogy with other systems at some level of abstraction, and including also rule-based representations and inferences based on these (cf. Kluwe and Haider 1990). In dealing with somehow "intelligent" systems, a third conceptual level may be invoked in addition to the physical-causal level and to the functional level at which one adopts a designer's viewpoint of the system, namely, the intentional level of description at which the system is being considered as if it were an independent intelligent being having goals, reasons for action, etc., of its own.

Even in the case of only moderately complex dynamic systems and certainly in the case of complex systems, the full causal and functional stories of a particular system normally are not available to the operator in the form of mental models. This means two things: first, that the mental system models even of operators dealing regularly with tasks such as explanation, troubleshooting, og prediction of system states under new circumstances are incomplete and may contain errors and contradictions; second, that such tasks are hard and may be impossible to carry out in realistic conditions (involving, e.g., time pressure during an emergency) even for trained operators. Again, the question is raised of how to support operators' formation of system models at appropriate levels through interface design, training, and manuals.

The introduction of the part-whole and means-ends hierarchies have emphasized that one and the same system can be *seen* (conceptualized, reasoned about) *as* very different things depending upon the viewpoint: as a material, spatio-temporal, causal physical system, as a functional whole created or evolved to serve some purpose, or even as an intentional system rationally responding to events in the environment. This point has been made in different ways in the cognitive science literature as applying equally to complex, man-made systems and to biological systems (Dennett 1981, 1987, Marr 1982, Pylyshyn 1983). One question for mental model theory to address is whether there is such a thing as one system mental model or whether an operator needs several different and distinct models and, if so, how they interrelate. In addition to system mental models, operator performance at the knowledge-based level involves the mental representation of yet more structure in order to deal with the full complexity of systems, interfaces, tasks, task domains, collaborators, organisations, and the environment at large. If the mental model concept proves useful, it seems that we may have to assume mental models of phenomena other than systems.

A "dynamic world model" may be called a "mental model" in some sense. The problems are that this "model" is hard to access through current methods and that it may be very different from knowledge representation at higher cognitive levels. The notion of mental models was tentatively introduced above as covering at least models of systems at the knowledge-based level, which means that the presumed role of the concept of mental models is at least to capture, at a deeper level than that at which knowledge-based performance has been considered so far, the knowledge representations and operations over them used in knowledge-based performance. At the rule-based level, mental models may be in operation simultaneously with rule-based processing, but it is not clear how knowledge-based mental models and rule-based processing are interrelated. We shall see some results later which suggest that mental models may in some cases include both declarative and procedural knowledge.

4. Current Theories of Mental Models.

Now let us turn to the discussion on mental models during the last decade or so. What is the rationale for considering mental models in the context of HCI or cognitive engineering ? The basic point of view has been expressed as follows: cognitive engineering tries to apply the relevant parts of what is known about the fundamental principles behind human action and performance to the design and construction of machines that are pleasant to use. Much is known in cognitive science that can be applied. On the other hand, our lack of knowledge is appalling (Norman 1986). Mental models research in the context of HCI provides a clear illustration of this situation. On the one hand, since the seminal work of Gentner and Stevens (1983), there has been a growing body of empirical studies and results on mental models in the HCI literature. It is felt by many that it would be difficult to explain many aspects of human behaviour without resorting to a construct such as mental models (Rouse and Morris 1986); or that, in HCI, one would like to know the nature of the mental models entertained by users in performing certain tasks through the use of computers as intermediaries in order to obtain constraints for optimisation of the computerized interface between task domain and user. These constraints may be used in improving aspects of the computer interface ranging from relatively low-level properties of displays like the use of menues, windows, or icons to facilities for, e.g., explanation and fault diagnosis (Wilson and Rutherford 1989, Moray 1987). On the other hand, there is no consensus on the nature of mental models and the range of phenomena they might be invoked to explain.

Two points are clear, however: (1) The study of mental models in the HCI context so far has been focused on the mental models of *users or operators* performing computerized and other tasks. (2) Research has focused on user's mental models of the *systems* they are operating. Mostly, such systems are physical systems of many different kinds, but they may also be intentional systems, as in the case of a "system" of waiting aircraft controlled by an operator on the airfield. In what follows, the expression "system mental models" will be used when needed. It has proved useful (cf. Norman 1983) to distinguish between:

(a) The target system. In a specific case, there may actually be two systems involved, one of which is normally considered the target system. Such cases occur when a computer is used as a tool for controlling a second system, as in process monitoring and control where the processing system, and not the computer, is considered the target system. The computer itself, apart from its system image (see below), is considered "invisible" or "transparent". Objective system models have been discussed in sect. 3 above and it turned out that an exhaustive model comprises a large, interrelated set of descriptions of the system at different levels of detail, from basic scientific descriptions of system components up through the means-ends hierarchy to descriptions of system functionality and purpose and in some cases including intentional system descriptions, descriptions of system elements at different levels of detail in the part-whole space, static and dynamic system descriptions, system descriptions in information processing terms, and I/O system descriptions. In practice, for all systems and users (including designers) no user has, can have, nor needs to possess an exhaustive system model. This is one basic reason why the study of mental models is important to HCI: had exhaustive system models been necessary to users *and* easy to acquire by users then that would have been the end of the story.

(b) The designer's conceptual model of the target system, i.e., the model which was used in the design of the system - if, indeed, any single model was used during the design of the system. This model is also used as a tool for understanding and teaching of the nature of the system.

(c) The target system image, i.e., the target system as presented to users and consisting of displays, structure, operation, and documentation and training contents. These elements together provide the descriptions of the system's operating principles presented to users as well as the feedback provided by the system. Ideally, the system image should be internally consistent, intelligible, functional, learnable, usable, and in consistent correspondence with the (designer's) conceptual model of the system. It is commonly assumed that the satisfaction of many of these requirements depends on the system image's being suitable to the mental models users are likely to form of the system. Furthermore, the system image itself plays a crucial role in determining users' formation of mental models. It is a common view that, due to the increasing "black box" nature of systems, the power and complexity of control, and the wealth of output information, the mental models that operators develop are increasingly in the hands of designers (Brehmer 1987, Hollnagel 1988).

(d) The user's mental model of the target system. This model reflects the user's understanding of the system. The purpose of a mental model is to allow the person to understand and to anticipate the behaviour of the target system in ways relevant to the tasks to be performed. For this to be the case, there has to be some kind of correspondence between the parameters and states of the mental model and the parameters and states of the system being modelled. Many other terms than "mental models" have been used about (d), for instance, "users' conceptual models" (Moran 1981, Young 1983); "conceptualisations" (Baggett and Ehrenfeucht 1988); "device models" (Kieras and Bovair 1984); "naive theories" (McCloskey 1983); and "folk theories" (Kempton 1986).

(e) The scientist's hypothetical conceptualisation of the user's mental model. Mental models are hypothetical constructs that assist scientists in understanding behaviour (Wickens 1984). The conceptual model should eventually include a model of the relevant human information processing and knowledge structures that make it possible for the person to acquire and to use a mental model to understand the system.

(f) The designer's conceptualisation of the user. This conceptualization should perhaps be cruder and have a somewhat different purpose than the scientist's user conceptualisation (Young 1983). Thus it may incorporate assumptions about the tasks to be performed and the methods to be used which invalidate it as a psychological model; or it may ignore detailed models of the actual mental processes engaged by operators and only incorporate higher-level structural models of the mental activities they can use. However, since the designer is largely responsible for the system image, and since the system image is becoming increasingly important in determining users' mental models, it is not clear to what extent the designer's conceptualization of the user should differ from that of the scientific observer.

A number of further items are sometimes descriptively necessary:

(g) The scientist's hypothetical conceptualization of the designer's conceptualization of the user's mental model. This theme may look convoluted, but too little is known about the user conceptualizations which are applied during the design process.

(h) The computer system's built-in model of the user if and when such a built-in model exists or is needed in order to facilitate interaction, as in user-adaptive tutoring systems or explanation systems.

Among items (a)-(h), it is specifically item (d) that we want to theoretically characterize through using the (system) mental models concept. So we take an (e) view on (d) in order to explain and predict user behaviour with the ultimate purpose of influencing (f), (b), and sometimes even (a). The reason for this is that the designer's model of the target system is a conceptual model which, ideally, should be based on the typical user's tasks, requirements, and capabilities including background, experience, power and limitations of information processing mechanisms, processing resources, and short-term memory limits.

Let us look at some attempts to explicitly define or at least characterize important aspects of system mental models.

System mental models have an important **causal** aspect to them, and some authors speak of "causal mental models" in order to emphasize the element of causality in the representation of these systems which enables people to provide a causal system account when, e.g., **explaining or predicting** system behaviour. It is commonly agreed that this type of mental models **can be mentally "run"** to some greater or lesser extent and that they are **qualitative** models rather than quantitative ones (cf. Brewer 1987). A mental model of a device specifies how the device works in terms of its internal structures and processes (Kieras and Bovair 1984). Rasmussen (1979, 1986, 1990) suggests that mental models have to do with structural knowledge or **representations of structural configurations of elements and their functional relationships** and are important to the **planning** of interaction with the environment in the knowledge-based domain. Any system's operation can be conceived in terms of **various levels of abstraction** so that a mental model of a system may be made up of mental models concerning physical form, physical function, functional structure, abstract function, and functional meaning (cf. sect. 3 above). The questions of how well connected or embedded such mental models can be and how useful it would be to consider the total set of mental models concerned with a single system as one mental model remain unanswered (Wilson and Rutherford 1989). Mental models represent **both structural and functional aspects** of an object or system and allow us to predict the consequences of contemplated actions (Marchionini 1989). Mental models are the mechanisms whereby humans are able to generate **descriptions** of system purpose (why it exists) and form (what it looks like), **explanations** of system functioning (how it operates) and observed system states (what it is doing), and **predictions** of future system states. Generation may be simply knowledge retrieval or may involve inference and reasoning (Rouse and Morris 1986).

In a recent study on mental models in HCI, the conclusion is that a mental model is a representation formed by the user of a system **and/or task**, **based on previous experience (including training and background knowledge) as well as current observation,** which provides most if not all of their subsequent system understanding and consequently **dictates the level of task performance** (Wilson and Rutherford 1989). It is also concluded that **mental models should be described in semantic terms rather than in syntactic terms** although, in the HCI literature, the formal mapping approach has often been associated with user's mental models. In this approach, formalisms are being used to describe the mappings between the user's task and the system's execution of the task in order to estimate the complexity of system use. Proposed notations include grammars (e.g. Moran 1981, Payne and Green 1986, Reisner 1984), production rules (Kieras and Polson 1985), and stacks (Card, Moran, and Newell 1983). The utility of these models as design aids and as psychological models has been questioned (e.g. Briggs 1988, Green, Schiele, and Payne 1988, Karat and Bennett 1989). Such syntactically based formalisms do not make sufficient contact with the user's semantic representations of the system and task. Formal mappings provide a very constrained and, with less than expert users, potentially misleading account of users' behaviour (Wilson and Rutherford 1989).

If we disregard syntactic theories, the views referred to above concur on a number of points: mental models are analytic, knowledge-based level semantic tools for system understanding. They cover both system form and system function at several levels of abstraction, and they cover both static and dynamic system understanding including system stucture and system processes. They are qualitative and can be mentally "run" to some extent. As such, correct system mental models are important or even essential to the performance of tasks which normally require systematic, analytical system knowledge such as system description, prediction of future system states, interpretation and explanation of system behaviour, and planning of actions on the system. The acquisition of system mental models is based on background knowledge as well as on all aspects of the system image, including training, instruction, and observation of feedback from system exploration.

This general explication of system mental models gives rise to a research agenda which is currently being addressed in the HCI literature. The agenda includes questions such as: How are mental models constructed or learned ? How are mental models used ? What are the building blocks of mental models ? How systematic and coherent are mental models ? Are mental models analogue or symbolic or both ? Do we have one or more mental models of a given system ? To what extent can mental models be "run" ? How can mental models be invoked in the identification of limits to human decision making and control performance and to the explanation of human error ? What is the relevance of mental models for non-analytical tasks such as normal system operation and the learning of normal system operation, or, in other words, how do mental models relate to rule-based and skill-based performance ? How do mental models work relate to work on task-action syntax ? How can mental models be modeled ? How can mental models be simulated ? How can knowledge of user's mental models be applied to system design in order to support human decision making and control performance ? Many of these questions will be addressed below.

All of these questions take the notion of mental models for granted. But we still lack firm theoretical foundations for doing so, that is, if the notion of mental models is to be more than a useful fiction in HCI research. So far, we have only looked at system mental models. But at least one pair of authors above (Wilson and Rutherford 1989) suggested that there may also be mental models of tasks and procedures (cf. Veldhuyzen and Stassen 1977, Jagacinxki and Miller 1978). Similarly, some express the view that mental models serve as assumptions that allow calculations of expected control performance (Rouse and Morris 1986). Rasmussen's "dynamical world model" is yet another relevant hypothesis in this context. So where does the application of the mental models notion stop or, how do we manage the proliferation of types of mental models ? And how does the notion of mental models relate to the many other types of mental representation of knowledge which have been proposed ? Significantly, proponents of the mental models notion take very different views on these issues, one extreme being the view that the notion should be pragmatically circumscribed more or less as was done above and may only have marginal utility to the explanation of human performance (Rouse and Morris 1986).

At the other extreme, one finds the view of cognitive psychologist Johnson-Laird (1983, 1989) which posits mental models as the central construct for the understanding of human knowledge representation and processing. Johnson-Laird's theory of mental models was developed in order to explain (1) context-dependent deductive reasoning and inference and (2) discourse comprehension in a way which (3) opposes the assumption of a formal mental logic for the purpose of explaining (1) and (2). Basically, explaining (1) and (2) amount to the same thing in this contaxt since the problem contents presented to subjects in the inference and reasoning experiments whose interpretation lies at the root of the theory are presented in discourse, and since discourse comprehension involves inference and reasoning.

At the core of the theory is a procedural semantics for the construction and manipulation of mental models on the basis of "propositional representations" which are themselves constructed from natural language input. In a slightly expanded formulation (Johnson-Laird's wording in boldface, cf. Bernsen 1991), the procedures involved include :

1. A procedure that begins the construction of a new mental model based on the propositional representation and its truth conditions (and introducing - sometimes arbitrary or default or prototypical - semantic structures by using the system's mental lexicon and knowledge representation of the domain as elicited by the discourse), **whenever an assertion makes no reference, either explicitly or implicitly, to any entity in the current model of discourse (**as established from scanning of this model)**.**

2. A procedure which, if at least one entity referred to in the assertion is represented in the current model (possibly determined through scanning of the model)**, adds the other entities, properties, or relations to the model in an appropriate way** (using the system's mental lexicon and knowledge representation of the domain as elicited by the discourse. This procedure may note (depending on the difficulty of this task and on how discriminatively the discourse is attended to) if there are (unused) alternative possibilities without necessarily specifying these completely. The system assumes by default that the speaker intends to communicate one consistent mental model and that its task therefore is to reconstruct this model on the basis of the speaker's communications).

3. A procedure that integrates two or more hitherto separate models if an assertion interrelates entities in them (by scanning models and using the system's mental lexicon and knowledge representation of the domain as elicited by the discourse).

4. A procedure which, if all the entities referred to in the assertion are represented in the current model, verifies (possibly through scanning of the model) **whether the asserted properties or relations hold in the model (and adds what has not been included so far - cf. 2 above).**

5. A revision procedure checking whether an assertion discovered (possibly through scanning of the model) **to be false of the current model can be rendered true by recursively modifying the model in a way consistent with the previous assertions. If not, then the assertion is inconsistent with the previous discourse.** (If it can, then the model is non-monotonically revised accordingly).

6. A (surely optional, mainly for solving the specific task of valid deductive inference) **revision procedure checking whether an assertion true of the current model can be rendered false by changing the model in a way consistent with the previous assertions. If not, then the assertion is a logical consequence of the previous assertions.**

7. A set of procedures which evaluate the truth or plausibility of the model (at least given non-fictitious discourse) with respect to world knowledge from perception or memory. In some cases, this evaluation may lead to renewed search for alternative interpretations of the discourse.

Johnson-Laird's theory assumes the existence of three different types of mental representation: propositional representations, whose status in the theory is unclear (Bernsen 1991), mental models, and images. Images correspond to views of mental models from a particular point of view, that is, images correspond to perception of the world and are based on mental models of the world akin to those 3-D, object-centered models constructed in the course of, stored from, and used in, visual perception (Marr 1982). Manipulations of mental models need not involve imagery, however. An important class of mental models are constructed from perception whereas others - which may represent physical relations as well as more abstract entities and relations - are "conceptual". Johnson-Laird (1983) offers a typology of mental models which, being highly preliminary, will not be discussed here.

So it appears that the notion mental models has an extremely general domain of application according to the theory. All knowledge of the world depends on the ability to construct mental models. Mental models are structural analogues of the world, that is, they have a structure which is analogous to the structure of states of affairs (objects, events, processes, actions, etc.) in the world as perceived or conceived. Mental models enable people to make inferences and predictions, to understand phenomena, to decide what action to take and to control its execution, and to experience events by proxy; and they allow language to be used to create representations comparable to those deriving from direct acquaintance with the world. There are no complete mental models for any empirical phenomena.

Interestingly, mental model theory does not include any distinction between the types of representation involved in skill-, rule-, and knowledge-based performance, respectively. Input-output rules for system behaviour, task-action models of system handling, and models used in diagnosing system malfunction would all seem to be considered as involving mental models. The conclusion seems to be that either we adopt a narrow mental models concept which may be arbitrarily circumscribed, or we adopt a very broad and possibly principled concept which, then, requires a number of internal distinctions to be made. Several authors recommend that HCI adopts the latter, more principled approach (e.g., Manktelow and Jones 1987, Wilson and Rutherford 1989). This of course raises the issue of how far one wants to go in HCI in dealing with the fine-grained details of human representations and algorithms, but no doubt one can be justified in going further than has been done so far.

A final point of relevance in this context is the following. Whereas mental models according to Johnson-Laird's theory are constructed on the fly during discourse comprehension, the mental models discussed in HCI are relatively permanent structures in long-term memory. We should therefore distinguish between (relatively) *permanent* mental models and *temporary* mental models. This distinction is somehow related to classical distinctions between semantic knowledge and episodic knowledge and between generic and specific knowledge. How fundamental is the distinction ? Discourse comprehension is not as such a learning process whereas the construction of, e.g., system mental models is a learning process. This learning process normally includes, i.a., discourse comprehension events. Since system mental models can be described and communicated, what starts out as a temporary mental model constructed through discourse comprehension may subsequently become part of a permanent model in long-term memory. Discourse comprehension draws upon a store of relatively permanent mental representations established through prior learning or otherwise and belonging to the subject's background knowledge and mental lexicon, in addition to drawing upon the subject's current experiences. The same is true of system mental model learning. The conclusions are that HCI is primarily interested in (relatively) permanent mental models and only secondarily in temporary mental models, and hence that we need a new, or revised, "procedural semantics" for the learning and manipulation (use) of permanent mental models as compared with Johnson-Laird's proposal for a procedural semantics for discourse comprehension. Also, those permanent mental models have to be compared with related notions such as frames, scripts, schemas, scenarios, naive or folk theories, etc. (see below). These two points are related. As Wilson and Rutherford (1989) observe, the mental model notion could easily be seen as theoretically redundant if it lacks computational ability. Given that the structures hypothesized to represent background knowledge, such as schemas or frames, are taken to have computational capability, it is the dynamic computational ability of a mental model beyond that presumed of background knowledge that provides the notion with its theoretical utility. There are several examples in HCI research of steps being taken towards a "procedural semantics" for mental models (e.g., Williams, Hollan, and Stevens 1983).

It should be noted at this point that, in addition to HCI mental models research and work in the Johnson-Laird paradigm and related paradigms, there is already available another, relevant body of empirical results and theory based on these. It comes from cognitive semantics (e.g. Lakoff 1987). This work can best be described as work on the structure and building blocks of human conceptual systems and their relationship to perception, on the one hand, and to language, on the other. The notion of cognitive models figures prominently in this tradition of research, which supplements the two lines of work previously mentioned in providing a deeper study of the typology of permanent representations used in the construction and manipulation of temporary mental models as well as in the learning and use of relatively permanent mental models of systems and tasks. In addition, the notions of typicality and prototypes are central to cognitive semantics. These notions can be expected to play an inportant role in a future, comprehensive theory of mental models.

5. Mental Models and Schemas.

Let the notion of a schema (Norman and Bobrow 1976, Rumelhart 1980) include related notions such as frames (Minsky 1975), scripts dealing with extended activities (Schank and Abelson 1977), scenarios (Sanford and Garrod 1981), and possibly also the entire set of semantical items included in the "mental lexicon". This is pure terminology, since the frame notion might just as well have been chosen as the more comprehensive one. The extension of the schema notion is unclear. It is tempting to say that (permanent) system and task mental models are simply classes of complex schemas. Some speak of a continuum here (Wilson and Rutherford 1989). Rumelhart (1984) describes a mental model as the total set of schemas instantiated at the time. Considered as theoretical notions referring to human knowledge representation and processing (and not as AI formalisms), schemas share the entire research agenda associated with permanent mental models above, including the question how schemas relate to skill-based and rule-based performance. If we ask for the specific questions concerning permanent mental models that research on schemas may illuminate, they include the "procedural semantics" for building mental models from prior schemas during learning, a theory of the use of mental models/schemas during work, and the typology of schemas as building blocks for these purposes. These questions still remain to be answered. Let us simply note a number of features of schemas.

Schemas are mental structures that underlie many aspects of human knowledge and skill and which have been used to account for a wide range of phenomena, such as the occurrences of inferences from "semantic memory" and other memory phenomena, natural language comprehension, perception, sensory-motor activity, and planning. Schemas can be conscious or unconscious, static and dynamic, and may include both procedural and declarative knowledge. Different cognitive domains have schemas with different structural properties. Many schemas may be representable in the way proposed by Minsky (1975) to represent frames, i.e., as hierarchical data structures representing stereotypical entities; distinguishing between more and less essential attributes of these entities; having slots defined by a set of variable default values which describe the nature of the information acceptable to each slot; filling slots with appropriate information through a matching process; and, because of the default values of the slots, making it possible to assume specific information even if this information has not been encoded at the time. A schema may become instantiated in competition with others as the one best fitting the data (thus acting as an interpretation of the data), and may become subsequently discarded in the light of further information which leads to the instantiation of an alternative schema. Schemas may be combined during comprehension to form more complex representational structures in ways that are currently not well understood. This is where a "procedural semantics" is needed. Johnson-Laird (1983) proposed that schemas provide the procedures from which mental models are constructed, e.g., on occasion of linguistic input. Manktelow and Jones (1987) use the properties of schemas including procedural knowledge to provide a deeper explanation of some of the experimental results which lie at the root of mental model theory, and present some arguments for the need of a procedural semantics of mental models in HCI.

Schemas interact with incoming episodic information to (1) modify the generic information in the schema, and (2) to produce instantiated schemas, i.e., specific cognitive structures that result from the interaction of old information from the generic schema and the new information from the episodic input. Thus, it is typical of schema research to study the nature and instantiation of one particular schema rather than, as in Johnson-Laird's procedural semantics for the construction of specific temporary mental models, to study the combination of several different schemas and possibly other knowledge structures (cf. Brewer 1987). It is not clear how deep this difference is since schemas are involved in both cases, but the focus of schema research is the focus one has to adopt when studying the use of an established system or task mental model during work, for instance, in "running" a causal mental model in a specific situation in order to solve a problem. A good heuristic for distinguishing between instantiated schemas and episodic models is to ask if an experimental subject knew the relevant global knowledge structure before coming into the experiment (Brewer1987). A factor which further complicates the situation and which has not been mentioned so far is the following. Both in the construction on-the-fly of a temporary mental model and in the use of a permanent mental model or schema in a particular situation, stored *episodic* information may be relevant and used. System troubleshooting, for instance, may be greatly facilitated if the operator happens to remember a similar case.

6. Methods for Studying Mental Models in HCI.

The methods for studying mental models in HCI are a large subset of those of cognitive science in general and the disciplines involved include most of those working within the framework of cognitive science. We would like to hypothetically identify user's mental models during learning, on-the-fly-construction, and during model application, and we would also like to be able to represent, model, and simulate user's mental models. So the methodologies used include analysis of verbal (interactive or non-interactive) protocols and think-aloud recordings, traditional cognitive psychology experiments and measurements for hypothesis-testing, developmental studies, expert-novice studies, simulation of possible psychological models and comparison of the results of that simulation with what humans do, field observation, comparison across cultures, comparison across time within the same culture, combined domain simulation and field observation in which an artificial domain is constructed that is of relevance to the realistic domains under consideration, linguistic analysis, analysis of succesful HCI systems, and, last but not least, conceptual studies of the mental models framework itself.

Some more specific points on methodology are the following:

At this early stage of mental models research, focus has been on system mental models in HCI and on mental models of linguistic input in the Johnson-Laird paradigm. Mental model simulations are still few. Representational formalisms are still not widely applied but include semantic networks with attached procedures, production rules, hierarchically structured concept diagrams, graphs, state transition tables, and other standard knowledge representation formalisms. Neural network approaches also exist, one of the original inspirations for the revival of neural network research during the 1980's being an attempt to approach the modeling of schemas (McClelland and Rumelhart 1986). Typically, and this is characteristic of current HCI research more generally, there is a gap between attempts to explicitly model mental models, their construction, evolution, and use, on the one hand, and the study of mental models in complex realistic domains, on the other. But the potential of the former, more fine-grained descriptions is at present limited to providing explanations of very elementary psychological phenomena (Rouse and Morris 1986). The latter is still mostly done using empirical methods and includes little explicit mental model representation and modeling. In the book by Gentner and Stevens (1983), most authors made the obvious choice of working with simple naturalistic domains for which there exist good explicit normative models, such as simple physical and mathematical systems and artefacts, naive mechanics, naive theories of heat, and naive theories of liquids. A link was drawn between research on mental models and research on naive or qualitative physics based on everyday experience (e.g., Hayes 1978, 1985, Forbus1983).

A further point which emerges here is that mental models can of course be studied, and lessons relevant to HCI learned, in domains other than those of computerized work. There are two reasons for this. The first is that we still need much more general knowledge about mental models. The second is that people's prior knowledge and mental models of domains other than computerized systems and tasks have important effects on their learning how to perform tasks on such systems (see below).

Concerning the difficulties in accessing mental models through empirical methods, Rouse and Morris (1986) make the following points:

- experimental methods provide at best only indirect insights into the form (e.g., spatial vs. verbal) and structure of mental models;

- access and manipulation of models may be confounded with perception of displays and controls and with response execution. This is an important point since, e.g., suboptimal explanatory or predictive performance may be due to either inappropriate and inadequate perception and interpretation of cues as to what the current state of the system is or to inadequate mental models. Inadequate cue utilisation may be due to, e.g., working memory overload during task performance, looking for the wrong cues, difficulties in precisely estimating system states from the cues available, difficulties in correctly combining cues, or to interference between the rule-based and the knowledge-based levels. Cue usage is often overlooked in discussions of mental models. The reason seems to be that we are still far from really working on a procedural semantics for mental model construction and use;

- only where users' observations and actions are very simple and hence can be assumed to be performed correctly, as in extremely simple control and prediction tasks, they are unlikely to interfere with mental model manipulation. This allows algorithmic modelling of input-output relations in users. From these relations, the structure and parameters of mental models can be inferred;

- in many control tasks, operators have to adapt to the task to perform acceptably. Here it is common to assume that mental models are perfect relative to the state equations of the system, and one simply has to perform an engineering analysis of the system to identify the model. But in many monitoring and control tasks, subjects have discretion in how to perform and may succeed with only imperfect mental models. This makes it is much more difficult to identify the "right" mental model used since the imperfections assumed in the analytical model may be different from the actual imperfections;

- verbalisations of a non-verbal (e.g., spatial or pictorial) image may result in severe distortions and biases;

- verbal reports mainly report on what people are thinking about and much less about how they are doing it. Experts notoriously have difficulty in verbalizing their expertise. Wickens (1984) notes that models for control are less able to be verbalised than models for detection and diagnosis.

From a meta-methodological standpoint, Wilson and Rutherford argue that, in much of the mental models literature, the chosen method has determined the form of the mental model reported, and the anticipated or desired form of the mental model has restricted or even determined the type of data collected and therefore the method used. In many studies on mental models, it is unlikely that the structure revealed through analysis reflects the user's complete mental model. They conclude that there is a need for a thorough exploration of methods for identifying shared mental models of technological and information processes; for the testing of these in a variety of circumstances of system and task, such as, e.g., the differences between interactive and noninteractive system tasks, the relative roles of contextual or external cues in these, and the implications for the formation and updating of operators' mental models; and for the development of strategies to derive design criteria from the models (Wilson and Rutherford 1989).

It is important to clearly understand the roots of these problems. We do not have an adequate taxonomy of systems, nor of system images. We do not have an adequate taxonomy of work domains. We do not have an adequate taxonomy of tasks. We do not have a common taxonomy of users very much beyond such distinctions as naive/novice/expert, and the like. We do not have a common framework in which to describe the components of the cognitive architecture of a user which are being addressed in a specific empirical study. And of course, we do not have an adequate taxonomy of mental models. The joint effect of all these lacks is that it is extremely difficult to analyse, interpret, and report on a mental model study in a way which both addresses the core issues and avoids overgeneralisations into domains/systems/system images/tasks/users where the results obtained are not strictly applicable. In the absence of the structures mentioned we need more careful analyses and presentations of empirical results. The suggestion here is to use a number of standard parameters for characterising empirical experiments which will facilitate their interpretation and analysis and prevent inappropriate conclusions to be drawn. The point is not the presentation but the thinking prior to the presentation. The following remarks are highly preliminary. Such parameters should include:

- methodological approach. This aspect is normally presented in a satisfactory manner;

- specification of the system image presented to subjects. The problems are (1) that it can be difficult to decide how comprehensively the system image should be reported since one often has do deal with a large amount of information. Selective reporting, however, may be wrongly selective and undoubtedly often is; (2) that far too little is presently known about the semantics of interfaces and the way this semantics is being handled by users' cognitive architecture;

- type of system(s) studied. Much improvement is needed. Normally, what is (selectively) presented is the individual system(s) which have been studied, but no attempt is being made to categorise the system with similar systems into a particular class. If this is not done, it becomes difficult to generalise in any principled way the results obtained. Is the system (relatively) complex or simple; of what type or category is it; is it real or simulated; if simulated, what idealizations have been made; is it general-purpose, such as a programming language, or is it highly specialized, such as a spreadsheet; is the relation between system, users, and tasks one of one system/many different tasks/ different classes of users, or is the relation one system/one task/one class of users, or ... ?

- type(s) of users involved in the study. This aspect is often presented in a way which is less than satisfactory. Users may be characterised as "novice" or "skilled", or they may be said to have "had some experience with" some very general class of systems to which the system under study belongs, such as "computer systems" in general. Given the nature of the empirical problem addressed, we may need much more knowledge about the individual subjects involved, in particular about their skills and their background knowledge, in order to be able to correctly interpret results;

- type of task(s) performed. Again, much improvement is needed. Typically, the individual tasks being performed by subjects are reasonably well described, but no attempt is being made to categorise the task(s) performed into a particular task category. Clearly, it makes a difference if the task studied is one of skilled process controllers controlling a system during normal operation or if the task is one of expert maintainers' doing deep troubleshooting on a system behaving abnormally in novel ways. Such differences will reflect on the nature of the generalisations one is allowed to make about mental models from the empirical results obtained. One would like to know, i.a., if the task is (relatively) well-defined or ill-defined, if it is general- purpose or highly specialized, if it is interactive or not, which kinds of activities the task, as studied, involves, etc. As Rouse and Morris (1986) point out, mental models of a system depend on the tasks to be performed. If the system is used in many ways (cf. the SKR framework) then multiple mental models may be developed. This means that a purely system-oriented and task-independent taxonomy is inadequate. A behaviour-oriented framework is also needed. The level of behavioural discretion in performing a task can range from being more or less full, as in physics problem solving and system design, to being almost zero, as in many process control tasks or in the use of assembly instructions where the choices to be made by subjects are few. Since the more task-dominated aspects of the operation of many engineering systems are becoming increasingly automated, human tasks are becoming increasingly discretionary. And the nature of model manipulation can range from being almost fully implicit, as in many process control tasks, to being highly explicit (verbal and/or pictorial), as in many cases of system troubleshooting. As concerns generalisability across domains, the specificity and the form of conceptualisations of mental models are limited by the location of a domain or task along the two dimensions presented. This again has implications for the potential usefulness of alternative identification methods for mental models.

7. Some Characteristics of System Mental Models.

We may illustrate the above methodological considerations as well as our state of knowledge with respect to system mental models by taking a closer look at some of the literature. Norman (1983) concludes, from more or less formal observations of non-expert users' handling of devices such as pocket calculators, computers, computer text editors, digital watches and cameras, video cameras and recorders, and the piloting of aircraft, that:

- system mental models are incomplete;

- people's abilities to "run" their models are severely limited;

- system mental models are unstable: people forget the details of the system they are using, especially when those details have not been used for some period;

- system mental models do not have firm boundaries: similar devices and operations get confused with one another;

- system mental models are "unscientific": people maintain "superstitious" behaviour patterns even when they know they are unneeded because they cost little in physical effort and save mental effort in situations where people are not entirely certain as to the working of the mechanism of interest;

- system mental models are parsimonious: often people do extra physical operations rather than the more demanding mental planning that would allow them to avoid those actions.

Most people's understanding of the devices they interact with is surprisingly meager, imprecisely specified, and full of inconsistencies, gaps, and idiosyncratic quirks. People's models contain only partial descriptions of operations and huge areas of uncertainties; people often feel uncertain of their own knowledge, even if 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.

The "running" of mental models has been studied by De Kleer and Brown (1983) using a very simple mechanistic system, a door buzzer. They distinguish between constructing a causal (system) mental model or qualitative simulation of the functioning of the device on the basis of the device's structure, components and their behaviour, *envisioning*, and the process of simulating the result of this construction to produce a specific behaviour for the device, or doing a straightforward simulation of the machine, by giving a chain of events each causally related to the previous one, *running*. While the study contains many observations pertinent to the more fine-grained analysis of mental models we are looking for, it has been criticized on the grounds, i.a., that qualitative simulations of this sort can be difficult or impossible in the case of complex systems (Kluwe and Haider 1990); and that the described bottom-up approach to the learning of system operations and system functionality is not cognitively plausible as it involves very complex reasoning and ignores the use of top-down information on system functionality and system component functionality. Use of such information makes it easier to figure out how individual system parts work (Rasmussen 1990).

When we move from considering non-expert users toward considering expert users, most students of mental models would intuitively assume that users' mental models of systems and tasks become increasingly comprehensive, detailed, and correct (cf. Katzeff 1990). Moray (1987), as a result of a study of learning process control through practice on a simple, simulated thermal hydraulic system, offers the intriguing suggestion that this is not the case. When operators become skilled they learn through induction from observing the displayed variables and system state transitions that the system can be decomposed into nearly independent subsystems, that some variables are systematically interdependent, and that some variables "dominate" others in the sense that changes in one variable can cause a change in the other, but not vice versa. As a result, the control operator's mental model of the system develops into a homomorphic decomposition of the real system, and is composed of what the operator believes to be quasi-independent subsystems of the real system. This homomorphic (many-to-one-mapping) model is incomplete, reduced, partial, and simplified as compared with an isomorphic (one-to-one-mapping) model of the system controlled. The advantages are simplicity and reduced mental work load. The disadvantage is that one cannot deduce the state of the total system from the states of the homomorph. There is no way back to an isomorphic system model from a homomorphic one. When the operator encounters something unusual, like a plant failure or a state which has not been encountered before, the homomorph is not sufficient to control the system. So, at the "end" of learning, after long practise, the operator has learned a repertoire of rules which suffices for all the situations likely to be experienced, and believes that his homomorphic mental model is isomorphic. And he behaves rigidly in **the** face of system failure or malfunction, trying to fit the data to his mental model rather than exploring alternative possibilities during fault diagnosis and management.

Does Moray's result contradict the standard intuition ? The process of induction performed through increased experience with the system is undoubtedly a very important one which needs further study. What happens if, as would often be the case, the operator during practice with the system experiences many kinds of system disturbance which he has to cope with ? This would certainly influence the inductive processes he performs, and it might turn out that, as a result, his mental model of the system would conform more to the standard intuition than to Moray's hypothesis. As Moray notes himself, the nature of the homomorph developed is determined by relevance to the task(s) for which the model is required. However, the notion of "conforming more to the standard intuition" is a vague one which might still leave ample space for the operators' development of homomorphisms, especially if Moray is right in assuming that humans have a "natural" (i.e., through the nature of inductive inference) inclination to develop homomorphisms. A further argument in support of this is that "complete" system models which can be mentally "run" to cover all cases of system behaviour, is an ideal fiction. So there is a strong need to discover the mechanisms operative in people's formation of the incomplete system models they actually have (on the induction of rule-based system models, cf. also Holland et al. 1986).

8. Mental Models and Learning.

Once again, this is a huge and underdeveloped domain. Users develop their mental models on the basis of their background knowledge and previous experience with related systems and tasks, their information processing capabilities and limitations, the tasks they have to perform with the system, and all aspects of the system image. Mental model development may reach a plateau once they have obtained a workable or functional result in terms of their performance criteria even though the mental model which has been developed is neither accurate nor comprehensive.

We want to know more about the answer**(s)** to the following question: Which kinds of teaching can and cannot be used for optimizing users' mental models relative to a given combination of system, tasks, and user ? Let us look at some results.

An important distinction is that between the teaching of theoretical system knowledge vs. the teaching of operational procedures for task performance. It has been a common assumption that users, operators, and maintainers should thoroughly understand the basic theoretical principles on which the design and operation of a system is based, such as the fundamentals of thermodynamics, heat transfer, fluid mechanics, solid mechanics, dynamics, electricity, and perhaps mathematics. However, there is little evidence that this training emphasis results in better and more useful mental models. Such emphasis may even degrade performance and is compatible with low performance in process control, electronics troubleshooting, and even mathematical problem solving. This means that knowledge organisation and guidance in the use of knowledge is just as important during training as is the content of knowledge imparted to trainees. The teaching of explicit, operational procedures for task performance tends to be as useful as the teaching of principles, and at least as useful as having both procedures and principles. A less extreme form of guidance involves informing trainees of how and when the knowledge gained during teaching should be used, without telling them exactly what to do. This form of teaching seems to be necessary if clues, analogies, and general principles are to be transferred succesfully to task performance. Such more or less procedural guidance cannot always be explicit since it cannot, e.g., anticipate all complex, unexpected, and novel situations (Rouse and Morris 1986).

The possibility of transfer of problem-solving skills to novel contexts may be increased if instruction is provided in multiple contexts, e.g., for more than one system (Rouse and Hunt 1984). It is not clear what the mental models and the operations performed on them look like that are able to produce this result.

The teaching of fundamental principles is one thing, the teaching of system models is quite another. Kieras and Bovair (1984) studied the nature of device models supporting novices' procedural learning in simple process control. One study compared two groups, one of which learned a set of (normal and malfunction) operating procedures for the device by rote, and the other successfully learned the device model before receiving the identical procedure training. Some of the procedures taught were inefficient. The model group learned the procedures faster, retained them more accurately, executed them faster, and simplified inefficient procedures far more often, than the rote group. It seems that the model group received "more meaningful" instruction material. The hypothesis proposed was that knowledge of how a system works helps by enabling the user to reconstruct by means of inference the operating procedures even if specific details of them have been forgotten. A second study demonstrated that the model group, through reasoning in terms of the model, was able to infer the procedures much more easily than the rote group which worked by trial and error on superficial relationships between the controls. A third study showed that the important content of the device model was the specific configuration of internal system components (system topology and power flow) and how they are related to the controls, and not the motivational aspects (the phantasy cover story), component how-it-works descriptions, metaphors, analogies, or general principles (e.g., general power flow in the system). The specific information is what is logically required to infer the procedures. Thus, the benefits of having a device model for process control depend on whether it supports direct and simple inference of the exact steps required to operate the device.

These results, it should be noted, do not contradict the widespread assumption that, for tasks such as system explanation, prediction, or troubleshooting, component how-it-works knowledge as well as metaphors, analogies, and knowledge of general principles can be essential. But the results are particularly interesting because they address a central issue in the discussion on mental models in the HCI context. As the authors observe, there are two opposing intuitions in the literature: (1) Having a system model is of great value in learning how to operate a system, or in being able to operate it once it is learned. (2) System models are unnecessary: we need not, e.g., know how the telephone system works in order to successfully use a telephone; or, in SKR terms, since a task such as process control can be performed with knowledge only of input-output rules for system behaviour and hence is a typical skill-based and rule-based activity, there is no need for system models in the training for process control tasks and for the many other HCI tasks which can be performed exclusively at the skill-based and rule-based levels. E.g., Young (1983) discussed mental models in a way which favours position (2). Kieras and Bovair support position (1) by pointing out that device models supporting inferences about exact and specific control actions are useful in supporting procedural learning. They also note, however, that the relevant how-it-works knowledge can be very superficial and incomplete, since the user does not need to have a full understanding of the system in order to be able to infer the procedures for operating it. If the device is very simple, or the procedures are easily learned, there may be no need to provide a device model. For some devices (e.g., the telephone), the user may be able to infer a usable device model without explicit instruction. Even if this used to be true, it is probably not true any more given the more complex facilities offered by modern telephone technology. Finally, if the user is taught a device model, but fails to learn it correctly, incorrect inferences may be made and performance may not be facilitated at all, or may be impaired. With risk of overgeneralisation, Kieras and Bovair's results lend support to the general conclusion that the provision of some kind of correct system mental model is useful to the efficient learning even of the most rule-based tasks such as the control and use of systems during normal operation. Kluwe and Haider (1990) make the related point that a representation of a system, though highly partial, fragmented, and incomplete, normally contains from the start elements of all the different kinds of system knowledge, i.e., knowledge-based and rule-based representations.

While Young's (1983) arguments in favour of the task-action grammar approach have not been fully confirmed by subsequent research (cf. sect. 4 above), his arguments against having blind faith in the usefulness of system mental models (which he calls "surrogate models") are worth reviewing here. Surrogate models are biased toward the knowledge-based aspects of system handling and are otherwise of limited applicability, e.g., for typical system control and operation tasks. The issue of learning is not addressed. Users quickly learn some familiar sequences of operations that short-circuit what problem solving they might have been doing using surrogate models. Even simple surrogate models are difficult to mentally manipulate (or "run") given simple problems. These models (in contrast to the Kieras/Bovair scenario above) say nothing about the structure of input sequences to systems. And they say little about possible user difficulties and errors: the surrogate model may be clean and efficient in predicting system output but has nothing to say as to whether its way of operation fits the user's task model or not. Hence they provide little information for the designer. In addition, most types of users do not need to know what happens inside the complicated systems they use. This last comment appears a little biased in the light of the current discussion. Young concludes that surrogate models focus exclusively on the device itself and take no account of the user, the task to be carried out, and their interrelationship, and do not provide help to the designer. A strong point about the task-action grammar approach is the emphasis on task models although these tend to be defined exclusively on the basis of the technological system at hand rather than taking into account users' previous task models derived from other, similar tasks, as well as users' already established models of the task domain. As Young convincingly shows using the example of different pocket calculators, users' established task and task domain models may sit very uncomfortably with the built-in task model of a given system. Task models and task domain models tend to be somewhat neglected in the study of system mental models.

Let us turn to the topic of prior knowledge and learning. People always approach the learning of new system/task combinations with prior knowledge and skills, including prior mental models, and this prior knowledge tends to affect learning. As said earlier, users' prior knowledge may to a large extent be regarded as consisting of "building blocks" in the form of schemas, which are applied both for the purpose of interpreting events, objects, and situations, and for actively responding to them. Prior knowledge can be usefully exploited by system images through, i.a., the use of metaphors and analogies which link up the new knowledge to be acquired with already existing permanent mental models. This principle has been widely applied in interface design although too little is known about the detailed mechanisms involved. Existing mental models can also be an impediment to instruction, however. Though consistent with everyday experience, they may be false with respect to the particular system to be learned, and they may not go away during training with the system. The result may be a combination of prior and new knowledge which may cause inadequate task performance. Taken together, these two points imply that instruction should take prior knowledge into account and try to remediate prior misconceptions as well as provide correct knowledge (Rouse and Morris 1986).

Katzeff (1990) studied the first phase of learning of text retrieval from simple instructions and from exploration of a three-level, full-text retrieval database. The topics of study were novices' learning of system mental models and the prediction of conflicts between prior knowledge, system image (clues to correct mental models), and necessary system knowledge. One reason why such conflicts arose with the result that subjects did not understand the system properly, was simply a lack of the appropriate schemas. If the appropriate schemas were there, failure could be due to the fact that the designer had not provided the clues sufficient to suggest adequate mental models of system behaviour. For an exploration of the system to succeed, clear, appropriate, and unambiguous system feedback is essential. Users will use all accessible hints (available in the system image) to form a model of system behaviour. Consistency in the hints provided is crucial. If clues are unclear, subjects are likely to use their schemas based on everyday life. The clash between the required mental models and the everyday life schemas will then be a source of difficulties. These schemas comprise task/system schemas or task/system mental models. Subjects come to a text retrieval systems with established procedures and expectations based on experience with text search in books and libraries, and it is essential to successful and rapid learning that a database system created to perform the same generic task either fits these procedures and expectations or makes explicit where they do not fit. Examples were also found where users developed a consistent mental model which, however, was different from the one intended by the system designer. Furthermore, users may persist in applying a false mental model despite the fact that the system does not respond as expected. Feedback is then interpreted to suit the false model.

Kempton (1986) studied non-technical users' own, implicit and naive theories or mental models of simple home heat thermostats which they routinely controlled during everyday life. The reason for using the term "theory" here is that users appeared to hold relatively systematic sets of beliefs about the functioning of the device under study involving abstractions, applying to many analogous situations, enabling predictions, and guiding behaviour (cf. McCloskey 1983). Two different and incompatible theories were identified: the feedback theory and the valve theory. It turned out that naive theories which are inconsistent with technical system knowledge (such as the valve theory, which likened the thermostat to, e.g., the accelerator of a car) may be highly functional in everyday practice. On some problems, they may lead to more optimal performance than technically more correct, but incomplete, theories (such as the naive feedback theory). This suggests that a theory that is useful for designing systems is not necessarily a good theory for using them. In some of the subjects studied, it was found that pieces of incompatible naive theories may be held by the same individual. The study confirmed that naive theories are resistent to change. If a change in the theory is seen to be needed, subjects may make non-principled, ad hoc repairs.

A different aspect of prior knowledge was illuminated by Baggett and Ehrenfeucht (1988) in a study on novices' construction of assembly toy objects from memorized instructions. It was found that procedural assembly instructions presenting a typical conceptualization of the object to be constructed yielded better structural and functional performance in the constructed object than did instructions presenting a minority conceptualization. The hypothesis proposed to explain this is that the typical conceptualization better matches subjects' "natural" conceptualizations of the task so that no interference between two sets of conceptualizations can arise. In addition, it became apparent that there may exist several different typical conceptualizations of the task depending on subjects' backgrounds (e.g., engineer vs. novice).

Gentner and Gentner (1983) studied naive subjects' deep inferencing based on different analogies of a target domain. An analogy conveys a partial overlap in relational systems between the objects of a source domain and the objects of a target domain, respectively, but conveys no particular overlap in the characteristics of the objects themselves. So, explanatory analogies are about systems of interconnected relations. The source domain and its interconnections is normally well understood by the subject using the analogy. This is why we can use analogies to help structure unfamiliar domains and to perform inferences in them. It was found that the use of different analogies leads to systematic differences in the patterns of correct and incorrect inferences in the target domain (electricity), because the different analogies provide different understanding of various parts of the target domain. This seems to imply that the subjects were actually thinking in terms of the specific analogy used (either electricity-as-water or electricity-as-a-mowing-crowd) to understand the complex system in question, rather than merely borrowing language from one domain as a convenient way of talking about another domain. Interestingly, it was also found that subject's source domain models (*in casu*, their model of water behaviour) may be incomplete, inaccurate, and even internally inconsistent. This is a well-established fact in naive physics where people's models often lead to the prediction of physically impossible sequences of events although these models seem to perform reasonably well for the predictive purposes of everyday life. Such models, too, may be inductively derived homomorphs (cf. Moray 1987).

Another question which has been discussed extensively is the question of how the mental models of novices and experts differ. Larkin (1983), in a study of problem solving in physics, argues that many differences in the problem solving performance of experts and novices can be related to the use of basically different problem representations or mental models. Novices use a "naive" mental model composed of familiar real-world objects and developed through operators that correspond to real time sequential developments. This gives a "runnable" model of the real problem situation. Experts are able to construct a "physical" representation or mental model that contains additional abstract entities such as forces and momenta and their related principles that are not familiar but have meaning only in the context of formal and quantitative physics. However, in both naive and physical representations the inferencing rules are qualitative, and not tied directly to equations. The physical representation produces new information in any order in a time-independent representation. This means that an expert is not confused by the actual objects and relations introduced in a problem, but focuses on the relevant abstract parameters to solve the problem. Expert problem solving is also characterised by the redundancy of the inferencing rules used. Experts fill slots in one or several pre-existing schemas to make a physical representation, starting with slots related to known quantities. Only when a schema is successfully instantiated to form a complete consistent representation is it accepted and translated into a mathematical representation. Since novices do not have such schemas, they have little guidance in selecting principles for application and have to use a primitive means-ends strategy.

It is highly likely that expertise involves a superior organisation of knowledge. It is less clear whether experts' mental models differ fundamentally from those of novices relating to the same subject-matter as Larkin suggests (cf. Johnson-Laird 1989) and as DiSessa (1983) contradicts favouring a more continuous view of the developmental process. Some argue that expertise involves a conceptual shift rather than just a refinement of the novice's perspective (Wiser and Carey 1983); many others, that expertise consists in highly developed repertoires of pattern-oriented representations. It seems clear that the shift from novice toward expert does not imply that all naive notions are discarded: naive, "pre-Newtonian" theories of motion were retained by students even after instruction in "correct" theories (DiSessa 1982, McCloskey 1983, Clement 1983). Such inconsistent models may create difficulties in novel situations. One could argue, though, that "real" experts do not have such inconsistencies in their knowledge bases.

9. Forms of Mental Models.

It is not surprising that there has been no consensus on this question so far, that is, if the notion of mental models is as comprehensive as Johnson-Laird suggested; if mental models are often built in part from prior mental models; and if we have interrelated mental models of tasks and systems combining declarative and procedural information. Only if we limit the notion of mental models to system mental models supporting a limited number of knowledge-based tasks might we presently consider offering something like a typology of mental models. But if we thus limit the notion, say, for pragmatic reasons, we may risk depriving it of a sound theoretical foundation. Within this limitation, Rasmussen's five-level taxonomy of mental models presented earlier (sect. 3) offers a useful way of thinking of some of the complexities of knowledge-based system mental models. At some of these five levels, several forms of representation are possible, such as differential equations, functional block diagrams, and "snapshots" (or sequences of these) of physical form (Rasmussen and Rouse 1981). Rasmussen's taxonomy offers a way to address some of the issues involved in the question: do users have one mental model of the system they are using or do they have several models ? The mental model of the system considered as a material, physical, and causal mechanism at various levels of detail, and the mental model of the system considered as a functional device at various levels represent different conceptualisations of one and the same system. When the system is being used as a tool to perform a certain task, a third perspective is introduced on the system. Focus is now on how to get the task done using the system rather than on internal system topology, composition, causality, or functionality. More work is needed to understand these inherently different system conceptualisations which affect the whole "set" of a subject toward the system including the way the system is perceived, and which, except for a purely theoretical view of the system for no particular purpose whatsoever, seem always to combine to some extent declarative and procedural (task related) aspects (cf. the discussion of structural, functional and distributed models in DiSessa 1986).

It should also be noted that, just like system descriptions, task descriptions can be made at various levels of generality and specificity. So, a "task mental model", say, of sending messages or of retrieving texts, can be characterized at various levels of abstraction. The existence of such different levels of abstraction in mental models of systems and tasks may be important to understanding creativity in human problem solving, including the use of analogy and metaphor. We need generic constructs to explain human behaviour in unfamiliar situations. Such situations will become more widespread in computerized work in the future, given that routine tasks are becoming increasingly automated.

A classical issue in discussions of mental models is the question: are mental models propositional or pictorial or both ? Johnson-Laird's (1983) proposal is that mental models are either physical or abstract-conceptual or both; that they are analogue in that they preserve the structure of the state of affairs represented; and that to a considerable extent they exploit our stored perceptual representations in doing so. Images, including conscious, pictorial representations, are views of mental models from particular perspectives. Cognitive semantics (Lakoff 1987) emphasizes the relationship between cognitive models and perception. There seems to be a significant consensus in HCI mental models research that mental models are frequently pictorial or image-like. Whitfield and Jackson (1982) found that air traffic controllers had difficulty verbalizing their "picture" of the state of the system controlled. Denis and Cartafan (1985) summarize a number of psychological findings demonstrating the productive use of perceptual representations for various cognitive tasks: 1. There is a superiority of pictures over words in memorisation such that lists of objects are learnt better if people are shown the objects than if they are shown words designating the objects. 2. There are positive effects of instructions to use imagery in learning word lists. 3. There are positive effects on verbal learning from the possession of special abilities for forming mental images. 4. There are positive effects of visualisation on spatial reasoning. 5. There are positive effects of visualisation on deductive reasoning. 6. There are positive effects of mental practise in learning motor skills. 7. A large majority of people exhibit angular distance-time regularities in mental rotation experiments. 8. A large majority of people exhibit scanning distance-time regularities in mental scanning experiments. 9. In a large majority of people, the time to check properties of objects is a function of the size of the mental image. Thus, 1-5 show that perceptual experience (1) and visualisation (2-5) are positive factors in learning and reasoning. (6) shows that mental practise is a positive factor in learning. (7-9) and in some sense also (2-6) show the existence of structural similarities between imagery and perception.

Forbus (1983) argues that a quantitative "analog" geometric representation simplifies reasoning about space. It does so by providing a simple method for answering a class of geometric questions. Qualitative spatial reasoning can be thought of as manipulating a set of symbolic descriptions of space, defined in terms of the underlying analog representation, i.e., those symbolically expressible spatial concepts are defined in terms of analog representations. People may find diagrams useful because they allow certain spatial questions to be decided by interpreting the results of perception. The marks in a diagram reflect the spatial relations between the things they represent, which allows us to use our visual apparatus to interpret these relationships as we would with real objects. Perception provides a simple decision procedure for a class of spatial questions. Qualitative spatial reasoning may involve a vocabulary of places (i.e., pieces of space such that all parts of it share some property) whose relationships are described in symbolic terms. This place vocabulary is embedded in a more quantitative, analog representation. The fact that people are willing to take the trouble of generating paper and pencil diagrams to interpret symbolically presented relational descriptions of space indicates that our fluency in dealing with space does not come solely from a set of axioms for reasoning with a relational description.

The study of pictorial aspects of mental models represents an important part of the HCI research agenda, for at least two reasons. One is that these aspects have to be taken into account in a "procedural semantics" for mental model construction and use. The second reason is that the spreading of graphical interfaces demands that the pictorial (and, in a wider perspective, the perceptual) aspects of user mental models support have to be studied. There seem to be vast possibilities for improving the support for mental model construction and use through graphical and other interface technology. The perspective is that we may approach the development of "direct" interfaces to systems in which computer tools literally make visible their operations and assumptions as well as the operations and assumptions of the systems that may be controlled through them. The more "direct" the mappings between physical (system) variables and psychological variables the better. Similarly, the user's qualitative feeling of control is enhanced if he perceives that manipulation is directly operating upon the objects of concern (cf. Norman 1986). These points ov view are central to current work on ecological interface design (e.g., Rasmussen and Vicente 1990).

10. Mental Models and Systems Design: Our Lack of Knowledge.

While the results and conclusions above can be used to provide some general orientations for systems design, the overall current status of HCI mental models research is summarized by Wilson and Rutherford (1989) as follows: it is difficult in the literature to find an example of an explicitly stated user mental model that could be used subsequently in systems design, and there is nothing expressed in terms of anything similar to effective procedures. There is a present no well established theory or model of what it is that operators learn about a system. The several different interpretations of the concept of mental models and its utility is a weakness. These differences are found between application domains but also at a more basic level between different professions: the man-machine systems community has used the concept in total independence of Johnson-Laird's usage of the notion. The notion of mental models is in danger of being shrouded with sufficient ambiguity to become all things to all people. Rouse and Morris (1986) note that there is a lack of solid empirical evidence in the literature to support or invalidate the many hypotheses proposed on mental models. The preceeding discussion may have illustrated why this is so. A "procedural semantics" for mental model construction and use clearly is a long-term project. But it may still be useful to keep this project or some related project in mind even in the shorter term and while preserving the **(v)**outlook necessary to HCI that emphasis in the study of mental models should be pragmatic and that HCI can make do with approximate results from its supporting disciplines.

References.

Anderson, J.R.: *The Architecture of Cognition.* Cambridge, MA: Harvard University Press 1983.

Baggett, P. and Ehrenfeucht, A.: Conceptualizing in assembly tasks. *Human Factors* 30, 269-84, 1988.

Bernsen, N.O.: Johnson-Laird's theory of mental models. Working Papers. Center for Cognitive Science. Roskilde University 1991 No.2.

Brehmer, B.: Development of mental models for decision in technological systems. In J. Rasmussen, K. Duncan, and J. Leplat (eds.): *New Technology and Human Error.* Chichester UK: Wiley 1987.

Brewer, W.F.: Schemas versus mental models in human memory. In P. Morris (ed.): *Modelling Cognition.* Chichester UK: Wiley 1987.

Briggs, P.: What we know and what we need to know: The user model versus the user's model in human-computer interaction. *Behaviour and Information Technology* 7, 1988.

Card, S., Moran, T.P., and Newell, A.: *The Psychology of Human-Computer Interaction.* Hillsdale NJ: Erlbaum 1983.

Clement, J.: A conceptual model discussed by Galileo and used intuitively by physics students. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

De Kleer, J. and Brown, J.S.: Assumptions and ambiguities in mechanistic mental models. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Denis, M. and Cartafan, M.: People's knowledge about images. *Cognition* 20, 49-60, 1985.

Dennett, D.C.: *Brainstorms.* Brighton: Harvester Press 1981.

Dennett, D.C.: *The Intentional Stance.* Cambridge MA: MIT Press 1987.

DiSessa, A.A.: Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science* 6, 37-75, 1982.

DiSessa, A.A.: Phenomenology and the evolution of intuition. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

DiSessa, A.A.: Models of computation. In Norman, D.A. and S. Draper (eds.): *User-Centered System Design.* Hillsdale NJ: Erlbaum 1986.

Forbus, K.D.: Qualitative reasoning about space and motion. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Gentner, D. and Gentner, D.R.: Flowing waters or teeming crowds: mental models of electricity. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Gentner, D. and A.L. Stevens (eds.): *Mental Models*. Hillsdale, NJ: Erlbaum 1983.

Gentner, D. and Stevens, A.L.**:** Introduction to Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Green, T.R.G., Schiele, F., and Payne, S.J.: Formalisable models of user knowledge in human-computer interaction. In T.R.G. Green, G.C. van der Veer, J-M. Hoc, and D. Murray (eds.): *Working with Computers: Theory versus Outcome*. New York: Academic Press 1988.

Hayes, P. J.: The naive physics manifesto.In D. Michie (ed.): *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press 1978.

Hayes, P. J.: The second naive physics manifesto. In J. Hobbs and R. Moore (eds.): *Formal Theories of the Commonsense World*. Hillsdale, NJ, Ablex 1985.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R.: *Induction.* Cambridge MA: MIT Press 1986.

Hollnagel, E.: Mental models and model mentality. In L.P. Goodstein, H.B. Andersen, and S.E. Olsen (eds.): *Tasks, errors, and mental models.* London: Taylor and Francis 1988.

Jagacinski, R.J. and Miller, R.A.: Describing the human operator's internal model of a dynamic system. *Human Factors* 20, 425-33, 1978.

Johnson-Laird, P.N.: *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, Cambridge University Press 1983.

Johnson-Laird, P.N.: Mental Models, in M.I.Posner (ed.): *Foundations of Cognitive Science*. Cambridge, MA, MIT Press 1989.

Karat, J. and Bennett, J.: Modeling the user interaction methods imposed by design. Paper presented at the Austrian Computer Society Conference on mental models and human-computer interaction. 1989.

Katzeff, C.: System demands on mental models for a fulltext database. *Int.J. Man-Machine Studies* 32, 483-509, 1990.

Kempton, W.: Two theories of home heat control. *Cognitive Science* 10, 75-90, 1986.

Kieras, D.E. and Bovair, S.: The role of a mental model in learning to operate a device. *Cognitive Science* 8, 255-73, 1984.

Kieras, D.E. and Polson, P.: An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies* 22, 1985.

Kluwe, R.H. and Haider, H.: Modelle zur internen Repräsentation komplexer technischer Systeme. Institute of Cognitive Science, Hamburg University 1990.

Lakoff, G.: *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press 1987.

Larkin, J.H.: The role of problem representation in physics. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

MacLean, A., Bellotti, V. and Young, R.: What Rationale is There in Design ? In Diaper, D., Gilmore, D., Cockton, G. and Shackel, B. (Eds.): *Proceedings of INTERACT '90: Third IFIP Conference on Human-Computer Interaction*. Amsterdam: Elsevier North-Holland 207-212.

Manktelow, K. and Jones, J.: Principles from the psychology of thinking and mental models. In M. Gardiner and B. Christie (eds.): *Applying Cognitive Psychology to User Interface Design.* Chichester UK: Wiley 1987.

Marchionini, G.: Making the transition from print to electronic encyclopaedias: adaptation of mental models. *Int. J. of Man-Machine Studies* 30, 591-618, 1989.

Marr, D.: *Vision: A Computational Investigation in the Human Representation of Visual Information.* San Francisco: Freeman 1982.

McClelland, J.L. and Rumelhart, D.E.: *Parallel Distributed Processing.* Cambridge MA: MIT Press 1986.

McCloskey, M.: Naive theories of motion. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Minsky, M.: A framework for representing knowledge. In P. Winston (ed.): *The Psychology of Computer Vision.* New York: McGraw-Hill 1975.

Moran, T.P.: The command language grammar: A representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies* 15, 1981.

Moray, N.: Intelligent aids, mental models, and the theory of machines. *International Journal of Man-Machine Studies* 27, 619-29, 1987.

Norman; D.A.: Some observations on mental models. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Norman, D.A.: Cognitive engineering. In Norman, D.A. and S. Draper (eds.): *User-Centered System Design.* Hillsdale NJ: Erlbaum 1986.

Norman, D.A. and Bobrow, D.G.: On the role of active memory processes in perception and cognition. In C.N. Cofer (ed.): *The Structure of Human Memory.* San Francisco: Freeman 1976.

Payne, S.J. and Green, T.R.G.: The structure of command languages: an experiment on task-action grammar. *Int. J. Man-Machine Studies* 30, 213-234, 1989.

Pylyshyn, Z.W.: *Computation and Cognition: Toward a Foundation for Cognitive Science.* Cambridge MA: MIT Press 1984.

Rasmussen, J.: On the Structure of Knowledge - A Morphology of Mental Models in a Man-Machine System Context. Tech.Rep.No.Risoe-M-2192. Risoe National Laboratory 1979.

Rasmussen, J.: Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-13, 257-66, 1983.

Rasmussen, J.: *Information Processing and Human Machine Interaction: An Approach to Cognitive Engineering.* New York: Elsevier 1986.

Rasmussen, J.: Mental models and the control of action in complex environments. In Ackermann, D. and Tauber, M.J. (eds.): *Mental Models and Human-Computer Interaction 1.* Amsterdam: North-Holland 1990.

Rasmussen, J. and Rouse, W.B. (eds.): *Human Detection and Diagnosis of System Failures*. New York: Plenum Press 1981.

Rasmussen, J. and Vicente, K.J.: Ecological interfaces: A technological imperative in high tech systems ? *International Journal of Human-Computer Interaction* 2,2,1990.

Reisner, P.: Formal grammar as a tool for analyzing ease of use: Some fundamental concepts. In J. Thomas and M. Schneider (eds.): *Human Factors in Computing Systems*. Norwood NJ: Ablex 1984.

Rouse, W.B. and Hunt, R.M.: Human problem solving in fault diagnosis tasks. In W.B. Rouse (ed.): *Advances in Man-Machine Systems Research* Vol. 1. Greenwich CT: JAI Press 1984.

Rouse, W.B. and Morris, N.M.: On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin* 100, 349-63, 1986.

Rumelhart, D.E.: Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer (eds.): *Theoretical Issues in Reading Comprehension.* Hillsdale NJ: Erlbaum 1980.

Rumelhart, D.E.: Understanding understanding. In J. Flood (ed.): *Understanding Reading Comprehension.* Newark NY: International Reading Association 1984.

Sanford, A.J. and Garrod, S.C.: *Understanding Written Language: Explorations of Comprehension beyond the Sentence.* Chichester UK: Wiley 1981.

Schank, R.C. and Abelson, R.P.: Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. New York: Academic Press 1977.

Veldhuyzen, W. and Stassen, H.G.: The internal model concept: An application to modelling human control of large ships. *Human Factors* 19, 367-80, 1977.

Whitfield, D. and Jackson, A.: The air traffic controller's "picture" as an example of a mental model. In G. Johannsen and J.E. Rijnsdorp (eds.): *Analysis, Design, and Evaluation of Man-Machine Systems*. London: Pergamon Press 1982.

Wickens, C.D.: *Engineering Psychology and Human Performance*. Colombus OH: Merrill 1984.

Williams, M.D., Hollan, J.D., and Stevens, A.L.: Human reasoning about a simple physical system. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Wilson and Rutherford 1989 Wilson, J.R., Rutherford, A.: Mental Models: Theory and Application in Human Factors. Human Factors, 1989, 31 (6), 617-34.

Wiser, M. and Carey, S.: When heat and temperature were one. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.

Young, R.M.: Surrogates and mappings: two kinds of conceptual models for interactive devices. In Gentner, D. and Stevens, A.L.: *Mental Models.* Hillsdale NJ: Erlbaum 1983.