# Mental Models and the Control of Actions in Complex Environments

**Rasmussen, Jens**

*Publication date:*
1987

*Document Version*
Publisher's PDF, also known as Version of record

Risø-M-2656

Risø-M-2656

# Mental Models and The Control of Actions in Complex Environments

Jens Rasmussen

Risø National Laboratory, DK-4000 Roskilde, Denmark
July 1987

RISO-M- 2656

**MENTAL MODELS AND THE CONTROL OF ACTIONS**

**IN COMPLEX ENVIRONMENTS**

Jens Rasmussen

ABSTRACT The concept of mental models has become an important ingredient in models of the cognitive control of human behaviour. The paper reviews different approaches to the definition of mental models taken in psychology and cognitive sciences, which typically have been considering selected aspects of human activities. The need for analysis of complex work scenarios is discussed, together with the necessity of considering several levels of cognitive control depending upon different kinds of internal representations. The development of mental representations during learning and adaptation to the requirements of a task is discussed. Finally, the role of means-end considerations in problem solving and in understanding of the functioning of purposive mechanisms is illustrated.

Julv 1987

Riso National Laboratory, DK 4000 Roskilde, Denmark

Invited Paper for the Austrian Computer Society;

6th Workshop on Informatics and Psychology:

"Mental Models and Human Computer-Interaction"

Schaerding, Austria, June 1987

ISBN 87-550-1336-8

ISSN 0418-6435

Grafisk Service Riso 1987

LIST *OF CONTENTS*

## INTRODUCTION

The concept of a mental model is widely discussed in studies of the interaction of humans with their environment. Unfortunately, the concept has become very ambiguous, having been adopted by researchers approaching human cognitive functions from very different points of view. The aim of the present paper is to discuss the concept as seen from the point of view of analysis and design of interfaces between humans and their work based on advanced information technology. What is the nature of humans' conception of their work content, and how can computer-based information systems be made transparent and support the proper mental models?

## THE CONCEPT

In general, the concept of mental model is used to characterise features of the resident knowledge base, representing properties of the task environment which can serve the planning of activities and the control of acts when instantiated and activated by observation of the actual state of affairs.

### Craik's Mental Models

An early attempt to characterise the notion of a 'mental model' was Craik's discussion (1943) of a mental model as a basis for explanation and understanding. Craik mentions three essential processes of reasoning: "(1) 'Translation, of external process into words, numbers or other symbols, (2) Arrival at other symbols by a process of reasoning', deduction, inference, etc., and (3) 'Retranslation' of these symbols into external processes (as building a bridge to a design) or at least recognition of the correspondence between these symbols and the external events (as in realising that a prediction is fulfilled)." He then argues: "A calculating machine, an anti-aircraft 'predictor, and a Kelvin's tidal predictor all show the same ability. In all these cases, the physical process which it is desired to predict is imitated by some mechanical device or model which is cheaper, or quicker, or more convenient in operation. .... By a model we thus mean any physical or chemical system which has a similar relation-structure to that of the process it imitates. By 'relation-structure' I do not mean some obscure non-physical entity which attends the model, but the f act that it is a physical working model which works in the same way as the process it parallels, in the aspects under consideration." He continues: "... in the case of our own nervous systems, the reason why I regard them as modelling the real process is that they permit trial of alternatives, in, e.g. bridge design, to proceed on a cheaper and smaller scale than if each bridge in turn were built and tried by sending a train over it, to see whether it was suf f iciently strong.,, Craik also emphasises the fact that models only represent a selected set of relations: "Any kind of a working model is, in a sense, an analogy. Being different it is bound to break down by showing properties not found in the process it imitates or by not possessing properties possessed by the process it imitates." (p. 51-53). The implications of the latter statement will be that for the representation of human knowledge in a complex working context more than one 'mental model' should be considered. In general, Craik's conception of a model is very close to that of the engineering profession.

## Two Recent Approaches to 'Mental Models'

After Craik's early use of the term model' for cognitive representations, the theme has been considered from two different points of view. Mental models are the bridge between the work environment to be controlled and the mental processes underlying this control. Consequently, a study can be approached by a study of human mental processes as well as by a study of work requirements, and these approaches result in different concepts. The approach from the psychological point of view quite naturally focuses on the explanation of human performance, which often will be influenced by the AI related cognitive science. The focus of this research will be on the nature and form of the mental model together with its role in human reasoning and its relations to the 'mind' . Consequently, the criterion of success will often be whether a theory can be phrased explicitly in procedural form for simulation on computer (Johnson-Laird, 1983).

In contrast, the approach based on studies of actual work performance and systems analysis will typically be looking for the content of the possible mental models which will be effective for a given task repertoire in a specific work domain. The criterion of validity of the theories in this case will be whether they are useful, i.e., whether functionality of human-machine systems can be adequately predicted.

The two approaches are supplementary rather than competing and interaction between them is important for the development of modern information technology. The psychologically oriented approach serves to identify human reasoning mechanisms and resource profiles, while the approach from analysis of work requirements is necessary so as to interrelate the different human mechanisms in a complex work situation. Together, such studies may add up to be a response to Brunswik's request for an ecological psychology as a basis for system design (Brunswik, 1957).

**The Cognitive Science Approach**

A typical representative of the cognitive science approach is Johnson-Laird (1983). He distinguishes "at least three types of mental representations: propositional representations which are strings of symbols that correspond to natural language, mental models which are structural analogues of the world, and images which are perceptual correlates of models from a particular point of view." Basically, he categorises the form of representations rather than their content - which is in correspondence with his discussion from the point of view of modelling the reasoning mechanisms. He distinguishes flat least two levels of representation: a representation of the sense of a discourse, and a representation of its significance (including what it refers to). This distinction is precisely what is captured in the theory of propositional representations and mental models.,' (p. 395)

Johnson-Laird derives his concepts of mental models from a discussion o.f basic reasoning procedures, e.g., syllogistic or propositional reasoning, hence the requirement of procedural, computational tests. He takes the position that "a natural mental model of discourse has a structure that corresponds directly to the structure of the state of affairs that the discourse describes". The structure Johnson-Laird has in mind appears to be the physical anatomy of the environment. He states on the nature of mental models: "Mental models owe their evolution to the perceptual ability in organisms with nervous systems." Referring to David Marr's (1982) work he concludes: "It is therefore safe to assume that a primary source of mental models – three dimensional kinematic models of the world - is perception.,' This formulation brings Johnson-Laird's concepts close to the concept of the "dynamic world model,, discussed in a subsequent section. Thus, Johnson-Laird's mental models basically are representations of the context of reasoning, the "background" (Searle, 1983). Johnson-Laird says: "The meaning of a sentence, according to the principle of compositionality, is a function of the meanings of its words and the syntactic relations between them. Meaning, however, is an abstract notion that reflects only what is determined by a knowledge of the language. The significance of an utterance goes beyond meaning because it depends on recovering referents and some minimal idea of the speaker's intentions. The truth conditions of the proposition expressed by a sentence therefore depend on the meaning of the sentence, its context of utterance (as represented in the current mental model), and the implicit inferences that it triggers from background knowledge." In this way, Johnson-Laird's mental models are representation of the context or background by which propositional reasoning- is possible, i.e., his 'mental models' include the intuitive knowledge which escapes the representation by the explicit formulation in terms of the semantic nets of the AI community.

In this respect, Johnson-Laird continues a well established tradition. The need for a representation of the context for understanding language as well as for control of skilled movements has long been discussed in the philosophical literature. Polanyi (1967) has made thorough studies of the importance of "tacit" knowledge. Mackie (1975) found it necessary to introduce the notion of a "field" as the representation of the context in his efforts to define causality in common sense descriptions of event sequences. Recently, Searle (1983) has argued the importance of the "background" which in his terms is the "non-representational" something underlying mental representations, "intentional states". Mental representations form a network of intentional states, and the semantic content of a state depends on its location in this network. However, "anyone who tries seriously to follow out the threads in the network will eventually reach a bedrock of mental capacities that do not themselves consist in intentional states (representations), but nonetheless form the preconditions for the functioning of intentional states. His arguments for the existence of the "background" is very analogous to the present arguments for the internal world model: "The background is necessary to account f or the f act that the literal meaning of a sentence is not a context-free notion, for understanding of metaphors, and to explain physical skills as for instance needed in expert skiing".

Other approaches to the mental model concept from the artificial intelligence point of view have been presented by Gentner and Stevens (1983). Typically, the contributions to this collection focus on the representation of reasoning procedures by means of computer programs and, consequently, the problem domains discussed are rather simple and derived from naive physics, or from the immediate experience of computer scientists with, for instance, calculators or word processors. In the present discussion, the tentative taxonomy of features of mental models discussed by Young (1983) is of interest. He mentions several classes of models, but focuses the discussion of models on <u>surrogates</u> (i.e., analogies) and cue-action <u>mappings</u> of students during the use of calculators. His conclusion is that surrogates are biased towards the "reasoning criteria", and suited for explanation and prediction while they are limited in the use of a device. Mappings are better suited to "provide an overall characterisation of the machine to orient the user's behaviour". These categories and their performance implications are closely related to declarative and procedural models underlying knowledge-based and rule-based performance discussed below.

A natural consequence of the computational approach is a focus on a well bounded aspect of mental representation in the form of a ,mental model, in order to be able to embed it in a computer program. This, in turn, means that only rather simple and well structured problems can be approached or the danger will be that the reasoning procedure will turn out rather artificial. A case in point is the efforts of Brown and De Kleer (see for instance 1983) to develop computational models of qualitative reasoning about the functioning of physical devices. In order to have a manageable model for simulation, they assume that explanation of the functioning is inferred bottom-up from a representation of the structure of the device in terms of the topology of component connections together with component properties. This approach can be useful to model 'naive physics, reasoning e.g., on simple kinematic problems such as predicting the movements of objects given the initial conditions. When considering, however, the functioning of rather complex goal oriented artefacts from a general question of "how the work", the resulting reasoning procedure becomes very artificial. We will return to this topic in a subsequent section discussing the use of means-end relations for functional reasoning.

### The Cognitive Engineering Approach

Approaching the question of mental representations from the point of view of actual work performance, a focus on selected, wellformed aspects of mental models cannot be maintained. From the analysis of verbal protocols from actual work (Rasmussen and Jensen, 1973) our attention was drawn towards the interaction of several different kinds of mental representations (Rasmussen, 1986). Furthermore, the analysis of human errors in process plant operation has shown that the interaction of performance under control of basically different kinds of mental representations has to be considered (Rasmussen, 1980). Other studies of human performance in actual industrial work situations also have shown the need for analysing the difference between the representations brought to work by persons with different professional backgrounds and levels of training (De Keyser, 1987; Hery, 1987).

A natural consequence of this approach will be that the mental representations under work situations cannot be formulated in one well structured simulation language. Several different research models of the various mental representations have to be accepted (see the 'catalogue of models, in Rasmussen, 1986), and validation of the models will basically be a test of their predictive ability for systems design, i.e., validation depends to a large degree on evaluation of system. during actual work conditions. In order to have a framework for mapping the properties of different kinds of mental representations, a discussion of the cognitive control of skilled work performance will be useful.

## MODES OF COGNITIVE CONTROL

The point of view of the following discussion, therefore, is taken from the analysis of performance in complex work situations and from the requirements met when designing computer based interfaces. For such purposes to be satisfied, it is necessary to study the interaction of a wide variety of mental strategies and models. In particular, study of the interaction and interference between different modes of cognitive control appear to be important for the understanding of erroneous performance.

When we distinguish categories of human behaviour according to basically different ways of representing the properties of a deterministic environment as a basis for control of actions, three typical levels of performance emerge: skill-, rule-, and knowledge-based performance. These levels and a simplified illustration of their interrelation are shown in Figure 1.

Skill-based behaviour represents sensori-motor perf ormance during acts or activities that, af ter a statement of an intention, take place without conscious control as smooth, automated, and highly integrated patterns of behaviour. In most skilled sensori-motor tasks, the body acts as a multivariable, continuous control system synchronising movements with the behaviour of the environment. This performance is based on feed-forward control and depends upon a very flexible and efficient dynamic world model. In some cases, performance is one continuous, integrated dynamic whole, such as bicycle riding or musical performance. In these cases the higherlevel control may take the form of conscious anticipation of upcoming demands in general terms, resulting in an updating of the state of the dynamic world model and thereby in the appropriate "modulation" of the skilled response. In general, human activities can be considered as a sequence of such skilled acts or activities composed for the actual occasion. The flexibility of skilled performance is due to the ability to compose from a large repertoire of automated subroutines the sets suited for specific purposes. The individual routines are activated and chained by perceived patterns that are acting as signs, the person is not consciously choosing among alternatives.

At the next level of rule-based behaviour, the composition of a sequence of subroutines in a familiar work situation is typically consciously controlled by a stored rule or procedure that may have been derived empirically during previous occasions, communicated from other persons' know-how as an instruction or a cookbook recipe, or it may be prepared on occasion by conscious problem solving and planning. The point is here that performance is goal-oriented, but structured by "feed-forward control" through a stored rule, in other words, the person is aware that alternative actions are possible and has to make a choice. The choice is based on 'signs' in the environment which have been found to be correlated to one of the alternative actions. Very often, the goal is not even explicitly formulated, but is found implicitly in the situation releasing the stored rules. The control is teleologic in the sense that the rule or control is selected from previous successful experiences. The control evolves by "survival of the fittest" rule.

In general, skill-based 'performance rolls along without conscious attention, and the actor will be unable to describe the information used to act. The higher-level rule-based co-ordination in general is based on explicit know-how, and the rules used can be reported by the person, although the cues releasing a rule may not be explicitly known.

During unfamiliar situations for which no know-how or rules for control are available from previous encounters, the control must move to a higher conceptual level, in which performance is goalcontrolled, and knowledge-based (knowledge is here taken in a rather restricted sense as possession of a conceptual, structural model or, in AI terminology, of deep knowledge. The level, therefore, might also be called model-based'). In this situation, the goal is explicitly formulated, based on an analysis of the environment and the overall aims of the person. Then a useful plan is developed - by selection. Different plans are considered and their effect tested against the goal, physically by trial and error, or conceptually by means of 'thought experiments'. At this level of functional reasoning, the internal structure of the system is explicitly represented by a "mental model" that may take several different forms. A very important aspect of the cognitive control to be captured by models of human behaviour is the dynamic interaction between the activities at the three levels.

## Skill, Rules and Knowledtqe in Problem Solving

Problem solving takes place when the reaction of the environment to possible human actions is not known from prior experience, but must be deduced by means of a mental representation of the ‚relational structure' of the environment. This structure must be represented symbolically in a mental model. A major task in knowledge-based problem solving is to transfer those properties of the environment which are related to the perceived problem to a proper symbolic representation. The information observed in the environment is then perceived as 'symbols, with reference to this mental model.

Formation of a proper representation depends on knowledge about the basic laws governing the behaviour of the environment. This phase of problem solving is finished when a representation in a framework familiar to the person is obtained - which means a mental model for which a set of rules for information processing is available. The representation then ceases to be an informative, symbolic framework and turns into a prescriptive system of signs that control the application of stereotyped process rules.

If the representation then is externalised in the form of a physical or graphic model, it is evident that the same kind of rule- and skill-based operation can be developed, as it is found for operation on a physical environment. The efficiency of formal, mathematical models and technical graphs and diagrams, as e.g., control engineers' Bode plots and pole-zero graphs, depends on the existence of a large repertoire of stereotyped manipulation rules used for solutions and predictions - often to a degree where the engineer's fundamental understanding of the conceptual basis has completely decayed.

The conclusion of this discussion is that patterns in a symbolic model configuration, as is the case with perceptual patterns of the physical environment, can act as signs. This is most clearly seen if externalised representations of the mental model are actually available in the form of physical models, e.g., an abacus for calculation, or in the form of graphs or other symbolic representations on paper or on visual information displays, forming artificial objects for manipulation. For display formats this means that rule- or skill-based control - "direct manipulation" - at a higher abstract level can be obtained if a symbolic display can be designed where there is a one-to-one mapping between the immediate appearance of the display and the properties of the process to be controlled. In this way, the same conceptual model may act as a symbolic representation when considered in relation to the elements of the environment and the laws controlling their relationships, and as a system of prescriptive signs when considered in relation to the rules for model transformation and data processing.

The distinction between signs and symbols in representations in the sense discussed above is equivalent to the distinction in semiotics and information science between prescriptive and informative texts (Morris, 1971; Eco, 1979). The important point in the present context is, however, related to the fact that the same text - or model - will be considered as either prescriptive or informative by the same person, depending upon the situation. An important question is therefore how the cognitive level needed is activated and what information in fact serves this activation.

The role of a functional representation as prescriptive signs has been studied from a semiotic point of view by Cuni and Boy6 (1981). They analysed the role of electrical circuit diagrams in terms of signs controlling activities during design, installation, and repair of electrical power supply systems in private houses, and investigated how different appearance of the same functional diagram was effective for support of the different activities.

## Skills, Rules and Knowledcre as Stages in LearninLcf a Skill

It is clear from the discussion in the previous section that the three levels of control are intimately interacting. In order to evaluate the degree to which the underlying models are separate concepts or just different aspects of the same internal representation, it may be useful to discuss how they relate to learning.

Distinctions between different categories of human behaviour similar to the SRK-levels have previously been proposed in relation to learning a skill. Fitts and Posner (1962) distinguish between three phases: the early or cognitive phase, the intermediate or associative phase, and the final or autonomous phase. If we consider that in real life a person will meet situations with a varying degree of training when performing his task depending on variations and disturbances, the correspondence with the three levels in the present context is clear.

In the three-level model, the final stage in adaptation to a task environment is the skill-based level. During training the necessary sensori-motor patterns develop, while the activity is controlled by other means. It may happen directly at the skilled level by imitation and trial-and-error, as for instance, learning to play an instrument by ear or children learning to talk, walk, etc. In other cases, control at the rule-based behavioural level will be efficient during development of the automated skill. The rules may be obtained from an instructor or a textbook, as is typically the case when learning to drive a car, to operate tools and technical devices supplied with an instruction manual, or to manage social interactions from "rules of good manners". And, finally, persons with a basic knowledge of the structure and functioning will be able to generate themselves a set of rules to control activities related to various purposes during early phases of learning. This involves what Anderson (1983) calls 'compiling declarative knowledge.,

Human errors are closely related to this learning process. Fine tuning of manual skills depends upon a continuous updating of the sensory-motor schemata to the time-space features of the task environment. If the optimisation criteria are speed and smoothness, the limits of acceptable adaptation can only be found by the once-in-a-while experience gained when crossing the precision tolerance limits, i.e. by the experience of errors or near-errors. Also at the more consciously controlled rule-following level, development of know-how and rules-of-thumb is depending upon a basic variability and opportunity for experiments to find shortcuts and identify convenient and reliable signs which make it possible to recognise recurrent conditions without analytical diagnosis. Involved in problem solving, test of hypothesis becomes an important need. It is typically expected that for instance process operators check their diagnostic hypothesis conceptually by thought experiments - before operations on the plant. This, however, appears to be an unrealistic assumption, since it may be tempting to test a hypothesis on the system itself in order to avoid the strain from reasoning in-a complex causal net.

An important point is that it is not the behavioural patterns of the higher levels that are becoming automated skills. Automated time-space behavioural patterns are developing while they are controlled and supervised by the higher level activities - which will eventually deteriorate - and their basis as knowledge and rules may deteriorate. In fact, the period when this is happening may lead to errors due to interference between a not fully developed sensorimotor skill and a gradually deteriorated rule system. This kind of interference is known also to highly skilled musicians when they occasionally start to analyse their performance during fast passages. It seems plausible also that this effect can play a role for pilots of about 100 hours flying experience, which is known to be an error-prone period among pilots.

Anderson (1983) also discusses the interaction between declarative knowledge and procedural knowledge. He describes the development of procedural knowledge during learning as a 'compilation'. Generally, compilation refers to a transformation of knowledge by means of a formal procedure. According to the SRK-framework, however, procedural knowledge derived by compilation of declarative mental models is a possible, but not an inevitable, first phase of rule-based behaviour. In later phases procedural knowledge is typically not derived from the basic, 'deep' knowledge, but has an empirical, heuristic basis, and compilation is not a suitable metaphor.

Following the lines of reasoning suggested above, the transfer of control to new mental representations is a very complex process involving change along several different orthogonal dimensions. First, when trained responses evolve, the structure of the underlying representation shifts from a set of separate component models toward a more holistic representation. This is discussed by Bartlett (1943) in relation to pilot fatigue, and Moray (1986) analyses how such model aggregation can lead process operators into trouble during plant disturbances, because the process is irreversible, i.e., the regeneration of a structured model needed for causal reasoning in unfamiliar situations is not possible from the aggregated model. The learning model implied in the SRK framework indicates that skill acquisition involves not only an aggregation of mental models. Typically, control by a structural, declarative model will also be replaced by an empirical procedural representation concurrent with a shift from a symbolic to a stereotype sign interpretation of observations. This means that training involves at least three concurrent and structurally independent shifts, in terms of aggregation, of declarative procedural knowledge, and of interpretation of information.

If this model of learning a skill is accepted and skill/rule-based performance is characteristic of professional activities in general, one would expect the basic causal or functional understanding to deteriorate. This is in fact what is the evidence found by Ackermann and Barbichon (1963) from their analysis of the Organisation of knowledge and the explanation of phenomena as presented by electrical and chemical technicians in industry. Based on an analysis of interviews, their conclusions were that the professional knowledge of the technicians was fragmented, showed a lack of relationship among phenomena, had barriers between theory, practice, and extra-professional life, and lack of relationship among various representations - mathematical, graphic, concrete, and analogical - of a particular phenomenon. One explanation could be, as the authors suggest, that theoretical knowledge is not used and that their findings reveal rudimentary memories from basic education. But, seen in our context, it could also be that basic symbolic knowledge and representations are typically used to support stereotyped lines of functional reasoning in various typical situations, and causal reasoning therefore turns into rule- and skill-based manipulations of symbolic representations which will thereby lose their symbolic nature and theoretical relationship. Analysis of the functional explanations of f ered by the technicians interviewed by Ackermann and Barbichon was characteristic by resort to what the authors call verbal nominalisms, i.e., the use of technical terms without understanding their content or relationship, or verbal logicism, as for instance pseudo-analogies or replacement of logical sequences with chronologic sequences from the problem context. A general trend found is the replacement of functional arguments by reference to human interaction, a tendency that can be explained by the tight relationships between human acts and symbols which are degenerated into mere signs.

## A TAXONOMY OF MENTAL REPRESENTATIONS

From this discussion, a taxonomy of mental representations can be proposed. It is proposed to restrict the term mental model to the representation of the relational structure' of the environment, to follow Craik's definition. This means, the mental model is a representation of the fundamental constraints determining the possible behaviour of the environment, i.e., it is useful to anticipate its response to acts or events when instantiated by state information. The study of errors have made it clear that a taxonomy of representations should not only consider the higher level cognitive functions related to inference and reasoning, the role of the body in the control of sensori-motor performance is an integrated part of the system. In the following section, the representations related to actual working performance according to the skill-,rule-,and knowledge distinction are discussed in more detail.

### Representations at the Skill-Based Level

Performance at the skill-based level depends on a dynamic world model which has a perceptual basis, like Johnson-Laird's mental model. The model is activated by patterns of sensory data acting as signs, and synchronised by spatio-temporal signals.

This dynamic world model can be seen as a structure of hierarchically defined representation of objects and their behaviour in a variety of familiar scenarios, i.e., their functional properties, what they can be used for, and their potential for interaction, or what can be done to them. These elements of a generic analogue simulation of the behaviour of the environment are updated and aligned according to the sensory information during interaction with an environment. The model is structured with reference to the space in which the person acts and is controlled by direct perception of the features of relevance to the person's immediate needs and goals.

This conception is similar to Minsky's (1975) "frames". The main and fundamental - difference is that Minsky's frames depend on a sequential scene analysis; they are structured as networks of nodes and relations, and they are basically static. Minsky defines frames as a data structure for representing stereotyped situations which are organised as a network of nodes and relations. Gibson's (1966) concepts related to "direct perception" are far more convincing, viewed as a model of the high-capacity information processing mechanisms underlying perception, sensori-motor performance in fast sequences, etc., than Minsky's symbolic information processing. The latter is more adequate for higherl-evel conscious information processing, i.e., the manifestations of the "dynamic world model" at the conscious level in terms of natural language representations.

The "dynamic world model,, in the present context is very similar to the mechanisms needed for the "atonement of the whole retino-neuro-muscular system to invariant information" (Gibson, 19@6, p. 262), which leads to the situation where "the centres of the nervous system, including the brain, resonate to information". This selective resonance relies on the existence of some kind of dynamic model of the environment. The control of skilled performance by an active analogue model raises some interesting modelling problems, in particular as seen from the point of view of digital computer based AI. Skilled behaviour cannot meaningfully be decomposed into parts without a shift in domain of description to neuro-physiology. Behaviour of an active model is not controlled by rules, but by the laws controlling the behaviour of the involved physical system or, as Searle has phrased it "there is no computational answer to that; it is just done by the hardware" (Searle, 1984). The role of the dynamic world model for representation of the context - 'background, - of higher level cognitive functions makes this a basic problem of simulating intelligent human behaviour by present AI technology, as mentioned already by Dreyfus (1972): the problem with computers is not that they lack a brain; but that they lack the human body. From this point of view it is worth noting the recent revival in terms of 'connectionism' (Feldman et al., 1986) or PDP, i.e., 'parallel distributed processing' (Rummelhart et al., 1986), of the interest in parallel processing and self-organising networks which was characteristic of the bio-technological research of the 60'es (see for instance Oestreicher et al. 1966).

To Gibson, perception is not based on processing of information contained in an array of sense data. Instead, the perceiver, being attuned to invariant information in space and time in his environment, samples this invariant information directly through all senses. That is, arrays of sense data are not stored or remembered. Instead the nerve system "resonates". In my terms, the world model, activated by the needs and goals of the individual, is updated and aligned by generic patterns in the sensed information, but the idea of an organism "tuning in" on generic time space properties is basically similar and leads to the view of humans as selective and active seekers of information at a high level of invariance in the environmental context.

Like the perceptual function, motor control is not based on stored behaviour patterns from prior encounters, but on a constructive process that generates the proper patterns on demand (Bernstein, 1967). This is demonstrated by the fact that the success of rapid movements is independent of the initial positions of limbs and that movements can be transferred to other metric proportions and limbs. This function must depend on schemata for generating complex movements with reference to the internal dynamic world model. An important ingredient in motor control is the dynamic feed-forward generation of patterns within this internal dynamic map which is updated and aligned by the sensory information.

From the role in human behaviour, some of the functional features of the world model can be summarised:

-It is able to control bodily movements in a feed-forward mode of control during fast sequences, i.e., it is capable of real time, quantitative, and precise simulation of the time-space patterns of the environment, and it is an active model.

-It is a hierarchical representation; it enables recognition of objects and scenes at the level of physical appearance; it makes it possible to identify objects by their functional values rather than their appearance; and patterns of purposive behaviour can be activated by high level intentions.

-There is a very efficient mapping between features of the environment and the model; i.e., a very efficient updating of the model is possible in response to changes, as well as easy transfer to "similar,, scenarios. This points to an analog model with elements representing objects and their functional properties and values, and consequently a one-to-one mapping of elements in the environment onto the model.

It is important to repeat that the three levels of control are not alternatives. The skill-based level is always active; the dynamic world model supplies the contextual basis for all performance (compare the role of Johnson-Laird's mental model), it directs attention, activates higher level performance, and based on higher level intentions expressed in terms of goals or activities it controls information gathering and transforms intentions, to control of movements.

## Representations at the Rule-Based Level

At the rule-based level, system properties are only implicitly represented in the empirical mapping of cue-patterns representing states of the environment and actions or activities relevant in the specific context supplied by the underlying dynamic world model. According to the definition adopted here, this representation does not qualify as a mental model since it does not support anticipation of responses to acts or events not previously met, and it will not support explanation or understanding except in the form of reference to prior experience.

In order to prepare for rule-based control of activities, however, conceptual relations may be important. Descriptive relations are useful in assigning attributes to categories and, therefore, to label scenarios and contexts for identification of items to retrieve from memory. Descriptive labels are the basis for updating of the focus for intuitive judgements and for establishing the proper "background" of action and communication. As could be expected from research on memory (Bartlett, 1932; Tulving, 1983), episodic relations are important for structuring of memory. Episodic relations appear to be important for labelling prototypical situations to serve as tacit "frames" or context for intuitive judgements and skilled performance.

## Representations at the Knowledge-Based Level

In the present context, the representations at the knowledge-based level constitutes the proper 'mental models, being representations of the relational structure of the causal environment and work content. Many different kinds of relationships are put to work during reasoning and inference at this level, depending on the circumstances, whether the task is to diagnose a new situation, to evaluate different aspect of possible goals, or to plan appropriate actions. Two kinds of relationships, i.e., part-whole and means-end relations appear to be particularly important for the specification of the content and direction of problem solving processes and will be considered in some detail. These two dimensions constitute the problem space. The part-whole dimension is well suited to delimit the section of the problem environment which is actually within the span of attention, whereas the meansend dimension specifies the level of generality at which the problem present will be considered, i.e., the language in terms of model concepts which is used. Figure 2 illustrates the trajectory of the changing focus of attention in this problem space of an engineer during fault-finding an a digital computer system.

In the functional means-end hierarchy, the functional properties of the environment are represented by concepts that belong to several levels of abstraction; see Figure 3. In the present discussion, the focus will be on mental models of a physical system built to serve some human goal. The lowest level of abstraction represents only the physical configuration of objects and their locations, the material configuration of the system. The next higher level represents the physical processes or functions of the various components and systems in a language related to their specific electrical, chemical, or mechanical properties. Above this, the functional properties are represented in more general concepts without reference to the physical process or equipment by which the functions are implemented, and so forth.

When moving from one level of abstraction to the next higher level, the change in system properties represented is not merely removal of details of information on the physical or material properties. More fundamentally, information is added on higher level principles governing the co-ordination of the various functions or elements at the lower level. In man-made systems these higher-level principles are naturally derived from the purpose of the system, i.e., from the reasons for the configurations at the level considered. A change of level of abstraction involves a shift in concepts and structure for representation as well as a change in the information suitable to characterise the state of the function or operation at the various levels of abstraction. Thus an observer asks different questions of the environment depending on the nature of the currently active internal representation.

In other words, models at low levels of abstraction are related to a specific physical world which can serve several purposes. Models at higher levels of abstraction are closely related to a specific purpose which can be met by several physical arrangements. Therefore shifts in the level of abstraction can be used to change the direction of paths, suitable for transfer of knowledge from previous cases and problems. At the two extreme levels of models, the directions of the paths available for transfer are in a way orthogonal, in that transfer at one level follows physical, material properties, while at the other it follows purpose.

At each level of abstraction reasoning depends on a particular type of model and rules for information processing. Therefore, shifting the level of modelling can be very effective in a problem situation because data processing at another level can be more convenient, the process rules can be simpler or better known, or results can be available from previous cases. A special instance of this strategy is the solution of a problem by analogy, which depends upon the condition that different physical systems have the same representation at higher levels of abstraction. Higher level models for one physical configuration may therefore be reinterpreted to solve problems related to a quite different, unfamiliar configuration.

Physical systems with known and invariant internal structure are responding to changes and to human acts according to basic laws of nature which therefore can be used to predict their behaviour. They are causal systems, and their response to physical changes for which no experience is available to an observer can be explained or predicted by means of bottom-up reasoning in the abstraction hierarchy, i.e., by functional analysis.

This approach is not possible for all the environments in which humans have to make decisions. Systems with a high degree of autonomous internal functioning, with self-organising and highly adaptive features (as for instance when humans are part of the work environment), will change their internal functional organisation continuously in order to meet the requirements of the environment and to suit their internal goals or performance criteria. Even though such systems basically may be controlled by laws of nature, their complexity in general makes it impossible to explain or predict their performance by functional analysis during real-life decision making. The alternative is to consider such systems as intentional systems controlled by motives or intentions together with the constraints on performance posed by the environment physically or in the form of conventions and legal requirements and by the limiting capabilities of their internal mechanisms.

Decision making in control of intentional systems is based on knowledge of the value structures of the system, the actual input from the environment of the system, and its internal, limiting properties - i.e., it is based on reasoning top-down @'in the abstraction with little or no consideration of the internal causal structures or functions. This is probably the reason why top-level executive decision makers, according to Mintzberg's study (Mintzberg, 1973), do not behave according to analytical decision models, but prefer live action and constant consumer contacts instead of analysis of abstract reports, and current information even gossip and hearsay - for statistics and status reports. Meeting people and considering hearsay is probably the best sources of information on current trends in value structures.

Many technical systems such as control systems and information processing systems are very complex and have no simple relationship between their basic physical processes and their function in the information domain. Therefore, predictions regarding their behaviour are more readily made when considering the systems as intentional systems (Dennett, 1971). Even in case of relatively simple systems, operators can be seen in verbal protocols to develop an explanation of system behaviour from a top-down "re-design" of a reasonable functional structure from its supposed purpose, rather than to collect information on its actual, physical structure.

An illustrative example of the role of means-end relations can be found when comparing a decision task which has to be performed in a one-level formal description with the performance when the intentional context is also available.

The difference may partly be due to the use of shifts in level of abstraction to find paths for transfer of solutions and strategies by analogy, but also due to support of memory and search for rules in terms 'of structures at other levels of abstraction. A good empirical piece of evidence is the reasoning experiment Wason and Johnson-Laird (1972). Their experiment showed significantly better performance when a problem was embedded in the subjects' every-day experience, compared to the same problem in an abstract formulation. The difference in performance in the two cases probably can be explained by the role of means-end relations in the actual problem solving. In the abstract formulation, the problem solving is based on formal, logical arguments at only one level of abstraction, on syllogistic logic which requires manipulation of abstract symbols and storage of intermediate results in short-term memory. Embedded in a familiar setting, the context defines an intentional system, in which the effects of the different decisions can very easily be inferred at the higher levels. The reasons for proper states can be inferred top-down. The problem is solved by top-down model modification, by transferring to a model of "reasonable states of affairs".

The role of a multilevel abstraction hierarchy in problem solving is most explicitly seen in Duncker's (1943) research on practical problem solving related to physical, causal systems (radioactive tumour treatment and functioning of a temperature-compensated pendulum). Based on verbal protocols, Duncker describes how subjects go from the problem to a solution by a sequence of consideration where the items proposed can be characterised by a "functional value,, feature pointing upwards to the problem, and a "by means of which" feature pointing downwards to the implementation of a solution; The relation to the means-end hierarchy is clear.

Yet another observation on the role of an abstraction hierarchy on understanding a mechanical device has been reported by Rubin (1920), who reports an analysis of his own efforts to understand the function of a mechanical shutter of a photographic camera. He finds that consideration of purpose or reason plays a major role in the course of arguments: he conceived all the elements of the shutter in the light of their function in the whole. He did not perceive the task to explain how the individual parts worked, but rather what their functions were in the whole. How they worked was immediately clear when their function was known. He mentions that he finds it an analytical task to identify the function of parts, the direction of thought being from overall purpose to the individual function (top-down considerations). The hypothesis necessary to control the direction is then readily available. This approach was found to have additional advantages: solutions of sub-problems have their place in the whole picture, and it is immediately possible to judge whether a solution is correct or not. In contrast, arguing from the parts to the "way they work" is much more difficult as a result of being a synthesis. Solutions of sub-problems must be remembered in isolation and their correctness is not immediately apparent.

An interesting, albeit indirect, demonstration of the importance of means-end relations for functional reasoning is the difficulties met by AI attempts to model the function of mechanical devices 'bottom-up' from the function of the components. De Kleer and Brown find that determining the function of a device like an electric buzzer solely from its structure and the behaviour of the parts require complex reasoning. The inference model proposed is based on an examination of the propagation of events through the structure. In an earlier presentation, a basic principle was the no-function-in-structure' assumption (Brown et al. 1981). In a later discussion, however, inference is guided by 'class-wide, assumptions and functional evidence, which in fact appear to be a representation of purpose in disguise. The resulting inference process appear to be very artificial, compared to the top-down inference process guided by functional considerations such as those described by Rubin. In the De Kleer-Brown model, it will be difficult to see the wood for trees, while Rubin's description appears to be guided by a birds' eye perspective.

A number of conceptual relations, in addition to part-whole and means-end relations discussed in the previous sections, will be useful for operation on a problem representation in the knowledge based domain. When means for action has been chosen from perceived means-end relations in a particular work context, causal relations are used to judge the effect of actions. Value aspects are important for choice and for assignment of priority in decision situations when the constraints given by goal specifications and functional requirements leave freedom for optimising consideration, as for instance related to cost, reliability, effort required, emotional aspects, etc. Choice among possible strategies in a work situation will depend on performance criteria, i.e., value aspect assigned to the work process, as well as its product. Generic relations define a concept as a member of a set or category in a classical Aristotelian classification, and can be used to label part of the environment and assign it to a category for which functional properties are readily available. The generic relations are, in particular, useful for drawing formal logical inference (syllogistic reasoning).

Based on this discussion, a summary of mental representations is presented in figure 4.

## CONCLUSION

In conclusion, for planning human-computer interaction during actual work it is necessary to consider the interaction between several modes of cognitive control which are based on quit different kinds of mental representations. More research is needed for identification of the content and structure of such mental representations during complex tasks. In particular, the interaction between mental processes across cognitive levels and

through time during actual work is important to be able to support decision making and work by modern information technology.

## REFERENCES


Ackermann, W., and G. Barbichon. (1963). Conduites intellectuelles et activite technique. Bull. CERP 12, (1), 1-16.

Anderson, J. R. (1983). The Architechture of Cognition. Cambridge, MA: Harward Univ. Press.

Barnett, J. A. (1982). Some issues of control in expert systems. Proceedings of the international Conference on Cybernetics and Society, Seattle, WA, October 28-30, 1982, 1-5.

Bartlett, F. (1932): Remembering. Cambridge Univ. Press.

Bartlett, F. (1958). Thinking,- An Experimental and Social Study. London: Unwin.

Bernstein, N. (1967).The Coordination and Regulation of Movement. New York: Pergamon Press.

Craik, K. J. W. (1943). The Nature of Explanation. Cambridge: Macmillan.

Cuny, X. (1971). L'Approche Psvcho-Semiologigue: etude d'un Code Gestuel de Travail. Paris: Laboratoire de Psychologie du Travail de 1'E.P.H.E. Report, November 1971.

Cuny, X. (1973). Les Commandements gestuels: Une experience avec ouvriers strangers debutants. Bull. Psychol. 307, (26), 14-16.

Cuny, X., and M. Boy6. (1981). Analyse semiologique et apprentissage des outils-signes: L'Apprentissage du schema dlelectricitd. communications, 33, 103-140.

De Keyser, V. (1987): Structuring of Knowledge of Operators in Continuous Processes: Case Study of a Continuous Casting Plant Start-Up. In: Rasmussen,J., Duncan,K. and Leplat,J. (Eds.): New Technology and Human Error, London: Wiley and Sons

De Kleer, J. and Brown, *J. S.* (1983): Assumptions and Ambiguities in Mechanistic Mental Models. In: Gentner, D. and Stevens, A. L. (Eds.): Mental Models. Hillsdale, N.J.: Lawrence Erlbaum.

Dennett, D. C. (1971). Intentional systems. J. Philos. 68, (4), February 25, 1971.

Dreyfus, H. L. (1972). What Computers Can't Do. New York: Harper and Row; New revised version, 1979.

Duncker, K. (1945). On problem solving. Psychological MonograDhs. Vol. 58, No. 5, Whole no. 270.

Eco, U. (1979). A Theory of Semiotics. Bloomington: Indiana University Press.

Feldman, J.A. and Ballard, D.H. (1986): Connectionist Models and Their Properties. Cognitive Science. Vol.6 No.3

Fitts, P. M., and M. I. Posner. (1962). Human Performance.

Monterey, CA: Brooks/Cole Publishing Co.

Gentner, D. and Stevens, A.L. (1983): Mental Models. Hillsdale, N.J.: Lawrence Erlbaum.

Gibson, J. J. (1966). The Senses Considered as Perceptual Systems. Boston, Ma: Houghton, Mifflin.

Herry, N. (1987): Errors in the Execution of Prescribed Instructions. Design of Process Control Work Aids. In: Rasmussen,J., Duncan,K. and Leplat,J. (Eds.): New Technology and Human Error, London: Wiley and Sons

Johnson-Laird, P. N. (1980). Mental models in cognitive science. Cognit. Sci.. 4, 71-115.

Johnson-Laird, P. N. (1983): Mental Models, Cambridge University Press.

Mackie,, J. L. (1975). Causes and conditions. in Causation and Conditionals. E. Sosa, ed. Oxford: Oxford University Press.

Minsky, M. (1975). A framework for representing knowledge, in The Psychology of Computer YiLsion. P. Winston, ed. New York: McGrawHill.

Mintzberg, H. (1973). The Nature of Managerial Work. New York: Harper and Row.

Moray, N. (1987): Intelligent Aids, Mental Models, and the Theory of Machine. Private communication. To be published.

Morris, C. (1971). aeneral Theory of Sicfns. Paris: Mouton.

Oestreicher, H.L. and Moore, D.L. (Eds.) (1986): Cybernetics Problems in Bionics. New York: Gordon and Breach.

Polanyi, M. (1958). Lersonal Knowledge. London: Routledge & Kegan Paul.

Polanyi, M. (1967).The Tacit Dimension. New York: Doubleday & Co.

Rasmussen, J. (1979). On the Structure of Knowledqe – A Morphology of Mental Models in a Man-Machine Context. Roskilde, Denmark: Riso National Laboratory, Report No. M-2192.

Rasmussen, J. (1980). What can be learned from human error reports, in Changes in Working Life. K. Duncan, M. Gruneberg, and D. Wallis, eds. London: John Wiley & Sons.

Rasmussen, J. (1986). Information processing and Man-machine Interaction. Elsevier Science Publishers New York:

Rasmussen, J. (1983). Skills, rules, knowledge: signals, signs, and symbols: and other distrinctions in human performance models. !EEE Trans. Syst. Man. And Cyb.. SMC-13, (3), 257 -267.

Rasmussen, J., and A. Jensen. (1974). Mental procedures in real life tasks: A case study of electronic trouble shooting. Ergnomics. 17, (3), 293-307.

Rubin, E. (1920). Vorteile der Zweckbetrachtung fur die Erkentnis. Z. Psychol. 85 210-233 Also in Experimenta Psychologica, 66-81. Copenhagen: Munksgaard.

Rumelhart, D. E., and D. A. Norman. (1981). Analogical processes in learning, in Cogiitive ikills. and 2heir Accru.isition. J. R. Anderson, ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rummelhart,D.E. and McClelland, J.L. (Eds.), (1986): Parallel Distributed Processing: Explorations in the Microstructure of Cogiition, Vol. 1 & 2. Cambridge: MIT Press.

Searle, J. R. (.1981). Minds, brains, and programs, in J. Haugeland, ed, Montgomery, Vermont: Bradford Books.

Searle, J. R.(1984): Minds, Brains and Science, 84 Reith Lectures.

Tulving, E. (1983): "Elements of episodic memory". Clarendon Press.

## LEVELS OF COGNITIVE CONTROL OF ACTIONS

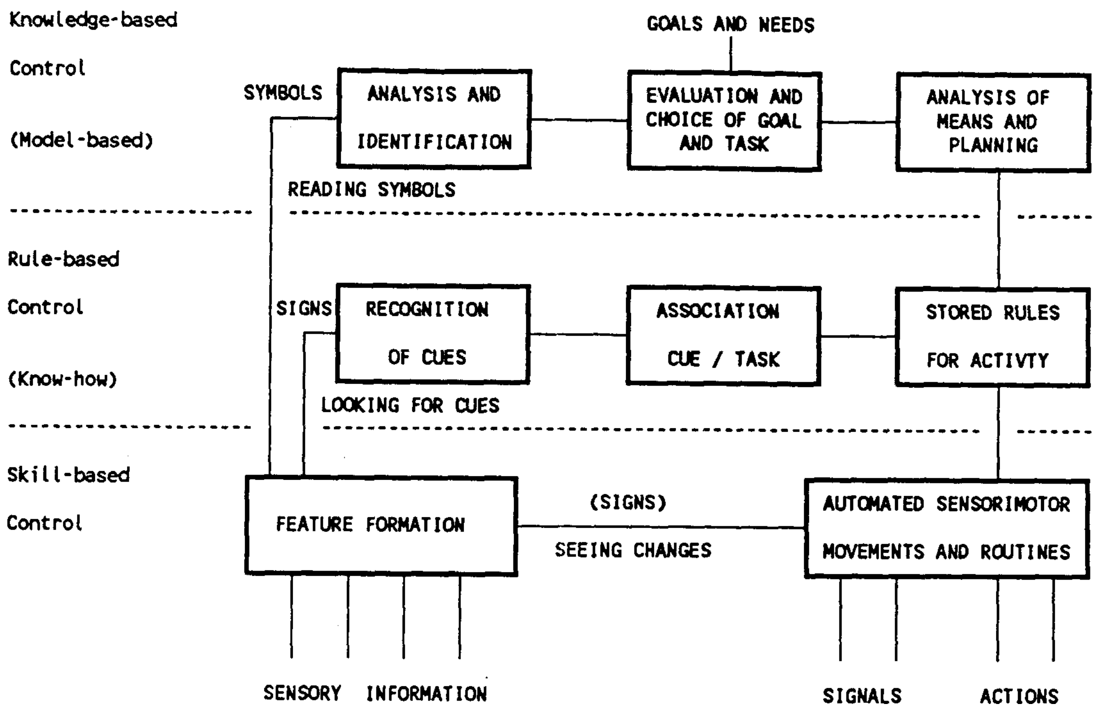


Figure 1. Schematic map illustrating different levels in cognitive control of human behaviour. The basic level represents the highly skilled sensori-motor performance controlled by automated patterns of movements. Sequences of such sub-routines will be controlled by stored rules, activated by signs. Problem solving in unfamiliar tasks will be based on conceptual models at the knowledge based level which serve to generate the necessary rules ad-hoc. The figure illustrates the flow of information, not the control of this flow. The figure is not meant to show humans as passive and subjects to information 'input'. On the contrary, they actively seek information, guided by their dynamic 'world model'.

**PROBLEM SPACE IN COMPUTER TROUBLE SHOOTING.**

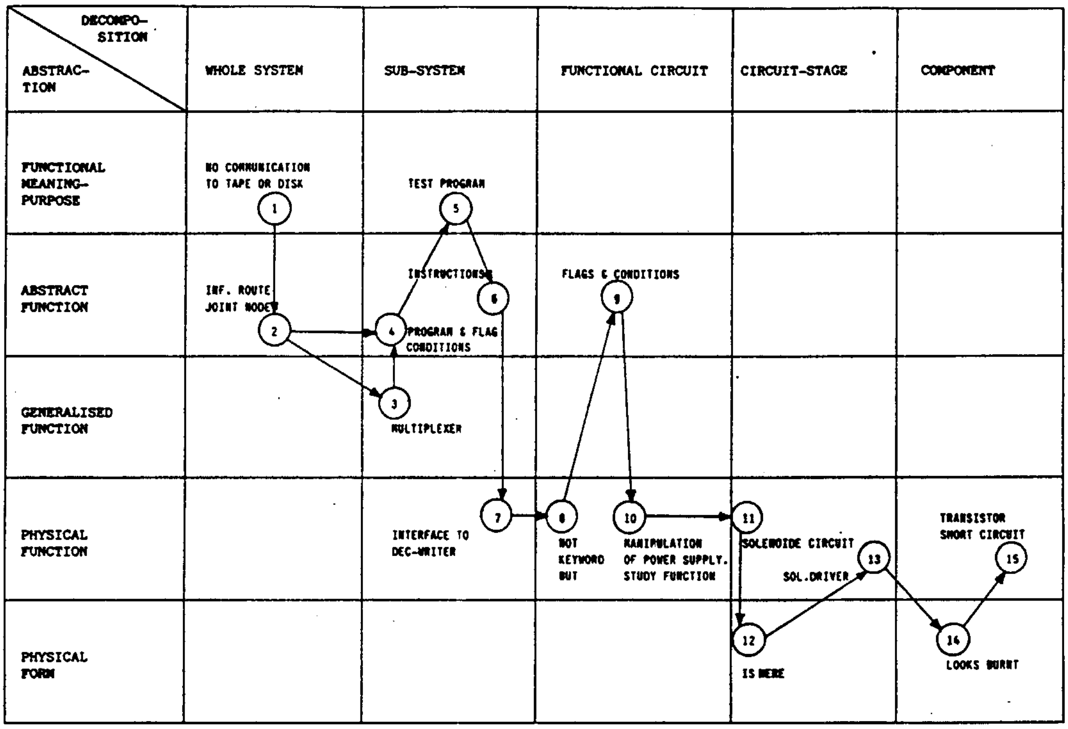


Figure 2. The problem space of computer trouble shooting can be illustrated by a map of two dimensions reflecting the level *s* of abstraction and decomposition considered in the individual statements of knowledge about the state of the system. Generally, in a resource-demand matching decision task, it will be expected that the decomposition is considered independently at each level of abstraction. In the present very selective task of locating a fault with reference to only one, normal, system state, a common decomposition at all levels with reference to the physical equipment was feasible for describing the trace found in verbal protocols. The figure illustrates the unstructured path of a specific case; each case will be different. Therefore, the process description is unsuited for a design basis model.

**MEANS-END ABSTRACTION HIERARCHY.**

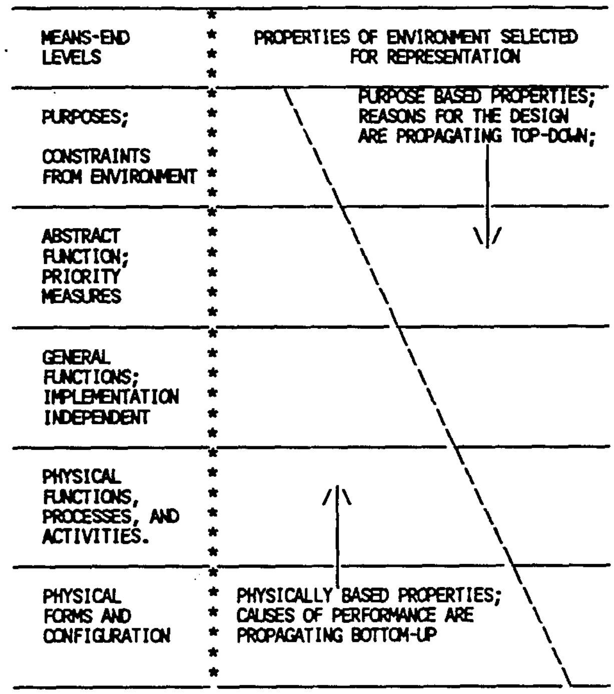


Figure 3. The functional properties of a physical system designed to serve human purposes can be described at several levels of abstraction. The figure illustrates that properties related to the purpose of the design, intentional relationships are predominant at the higher levels, while properties derived from physical properties are predominant at lower levels.

**SUMMARY OF THE BASIS FOR COGNITIVE CONTROL**

| BEHAVIOUR | REPRESENTATION OF PROBLEM SPACE | PROCESS-RULES |
|---|---|---|
| KNOWLEDGE-BASED | MENTAL MODEL; EXPLICIT REPRESENTATION OF RELATIONAL STRUCTURES; PART-WHOLE, MEANS-END, CAUSAL, GENERIC, EPISODIC, ETC. RELATION | HEURISTICS AND RULES FOR MODEL CREATION AND TRANSFORMATION; MAPPING BETWEEN ABSTRACTION LEVELS; HEURISTICS FOR THOUGHT EXPERIMENTS |
| RULE-BASED | IMPLICIT IN TERMS OF CUE-ACTION MAPPING; BLACK-BOX ACTION-RESPONSE MODELS; | SITUATION RELATED RULES FOR OPERATION ON THE TASK ENVIRONMENT, I.E., ON ITS PHYSICAL OR SYMBOLIC OBJECTS; |
| SKILL-BASED | INTERNAL, DYNAMIC WORLD MODEL REPRESENTING THE BEHAVIOUR OF THE ENVIRONMENT AND THE BODY IN REAL TIME | NOT RELEVANT - AN ACTIVE SIMULATION MODEL IS CONTROLLED BY LAWS OF NATURE, NOT BY RULES; |

Figure 4. Schematic illustration of the representations of the regularities behind the behaviour of the environment which are used for control of behaviour at the different levels of cognitive control.

**Risø National Laboratory** **Risø - M - 2656**

| Title and author(s) | Date July 1987 |
|---|---|
| MENTAL MODELS AND THE CONTROL OF ACTIONS IN COMPLEX ENVIRONMENTS<br><br>Jens Rasmussen | Department or group |
| | Groups own registration number(s) |
| | Project/contract no. |
| Pages 40 Tables Illustrations 4 References 42 | ISBN 87-550-1336-8 |

**Abstract** (Max. 2000 char.)

The concept of mental models has become an important ingredient in models of the cognitive control of human behaviour. The paper reviews different approaches to the definition of mental models taken in psychology and cognitive sciences, which typically have been considering selected aspects of human activities. The need for analysis of complex work scenarios is discussed, together with the necessity of considering several levels of cognitive control depending upon different kinds of internal representations. The development of mental representations during learning and adaptation to the requirements of a task is discussed. Finally, the role of means-end considerations in problem solving and in understanding of the functioning of purposive mechanisms is illustrated.

**Descriptors**

**Available on request from Risø Library, Risø National Laboratory, (Risø Bibliotek, Forskningscenter Risø), P.O. Box 49, DK-4000 Roskilde, Denmark. Telephone 02 37 12 12, ext. 2262. Telex: 43116, Telefax: 02 36 06 09**

RR950000 99

**Available on request from.**
**Risø Library,**
**Risø National Laboratory, P.O. Box 49,**
**DK-4000 Roskilde, Denmark**
**Phone (02) 37 12 12 ext.2262**

ISBN 87-550-1336-8
ISSN 0418-6435