Philosophica 62 (1998, 2) pp. 9-31

REVISITING THE MENTAL MODELS THEORY IN TERMS OF COMPUTATIONAL MODELS BASED ON CONSTRUCTIVE INDUCTION

Stefania Bandini, Gaetano A. Lanzarone & Alessandra Valpiani

1. Introduction

In the theoretical framework developed by Johnson-Laird (1983; 1993) about mental models and their role in human reasoning, *induction* is defined as a thought process starting with a set of observations, whose goal is to frame a hypothesis that reaches a better description or understanding of this information in relation to a background of general knowledge; any hypothesis can be evaluated and as a result, maintained, modified or abandoned. A distinctive feature of induction is that its conclusions increase knowledge in a plausible way.

The Johnson-Laird theory describes how such a process generates hypotheses through procedures of construction, manipulation and revision of mental models; it does not, however, give an explicit and formal (i.e., useful in the design of a computational model) description about how conclusions can go beyond the specific events considered to enrich the initial information.

The reinterpretation and development of this part of the Johnson-Laird theory, within the framework of a computational system, is the main goal of the work here presented, as the construction of such a system requires the adoption or development of a more articulated investigation into the necessary model of inductive process. The topic, the methods and the theoretical framework here used for leading the investigation belong to Cognitive Science, focusing on the study of conceptual systems and mental operations and structures through use of several disciplinary approaches and results (in our case, from Artificial Intel-

ligence to Logic Programming, passing through Cognitive Psychology and Qualitative Physics). Within this scenario, the main points of the work described in this paper are:
(1) adding a computational perspective to the conceptual system of mental models involving induction by introducing some recent advances in Machine Learning, namely, *constructive induction* (Idestam-Almquist, 1992);
(2) introducing a model and a computer system for representing and manipulating qualitative and common sense knowledge and reasoning about the physical world within the conceptual framework of mental models (Carassa et al., 1995);
(3) developing a system architecture for simulating the basic operations of mental models theory (*construction, manipulation, revision*) and of the induction process (*incremental learning by analysed examples*).

With respect to the first point, the generation of hypotheses as a succession of mental models is considered to be the preliminary stage of the induction process, while constructive induction is the logical operation used for selecting information and evaluating hypotheses.

A computational solution for formulating general hypotheses and enriching knowledge, coherently with the general mental models theory, covers the non-defined parts part of the Johnson-Laird theory. Two basic assumptions are needed:

- the set of hypotheses is a succession of examples, following the conceptual structure that defines a mental model as an example of the specific situation that it represents;
- the evaluation of mental models is a *constructive induction* problem, i.e., through the analysis of hypotheses, new knowledge with respect to some background knowledge is produced and integrated in a plausible way.

Regarding the second point, the circumscribed knowledge topic is *qualitative common sense reasoning about physical systems* (Hayes, 1979), a research area thoroughly investigated in Artificial Intelligence (Weld and De Kleer, 1990), which involves in an interdisciplinary way other crucial research topics in Cognitive Science, such as studies about cognitive physical pattern preservation in education (DiSessa, 1982), or specifically about mental models (Carassa et al., 1995) (Forbus and Gentner, 1997).

Following Holland et al. (1986), we are in a problem-solving context, i.e. to yield inferences that increase background knowledge in order to construct explanations of a physical situation, inductively drawing meaningful and holding relations (e.g., causal). When initial premises describing a circumscribed physical situation are given to the system, it tries to reconstruct a physical process that is able to generate the situation, on the basis of a given physical model contained in the knowledge base of the system (structured as a process-based knowledge representation, derived from the qualitative process theory - QPT (Forbus, 1984)). With this aim and by means of experimental manipulation, the system runs and stores a set of tentative explanatory models (examples), each corresponding to a hypothesis about the relations and the dynamic aspects of the physical process. More specifically, mental models of the physical world are modelled by means of an extension and a logic programming interpretation (Bandini et al., 1988) of the above mentioned QPT, whose main constructs allow mental models theory components to be described and manipulated without upsetting the structure of the theory.

Finally, with respect to the third point, a specific architecture has been designed and developed to allow simulation of the entire process (construction, manipulation, revision). Specifically and from a functional point of view, when initial premises describing a physical situation are given to the system, this tries to *reconstruct* a physical process able to generate the situation (as above mentioned). By means of experimental *manipulation*, it creates a series of tentative explanatory models, each corresponding to a hypothesis about the causes and the dynamics of the involved physical process. When an anomaly on parameters is encountered that the background knowledge is unable to justify, the search for an explanation by means of a *revision* process (Johnson-Laird, 1983) leads to building a set of alternative models. New models (as examples) are produced by adding new entities or changing qualitative parameters of already represented entities. Thus, the repeated manipulations constrained by previous knowledge and the evaluation in relation to the premises allow the system to draw a plausible explanation about the causes and the dynamics of the physical process generating the situation.

The construction of a model, which is consistent with both the background knowledge represented in the system and the examples about the considered situation, leads to a plausible explanatory hypothesis. Such a hypothesis is specific in that it is valid only for the particular situation described by the initial observations. Moreover, the system may yield overall hypotheses through comparative analysis of the incrementally generated models. These hypotheses cover classes of situations and enhance the background knowledge of the system, allowing *incremental learning by analysed examples* to be performed, guided by constructive induction inference.

The structure of the paper is the following. Section 2 introduces the main aspects of the mental models that we implemented in the system. Section 3 describes how mental models of physical phenomena are built and their temporal evolution simulated. Section 4 shows how the search for an explanation leads to building alternative models; then it describes how comparative analysis of models can be seen as a constructive induction problem and thus general conclusions can be induced. Finally, some concluding remarks are made.

2. Mental Models and common sense reasoning about the physical world

Common sense and qualitative reasoning about physical systems, within the general framework of Artificial Intelligence and from a Cognitive Science perspective, implies the computer manipulation of physical models not entirely derived from or described by the quantitative classical models developed in Physics, but involves mental operations and structures that humans use when reasoning about the physical world. For this reason, representational languages, computational environments and theoretical models have been developed over time, producing meaningful contributions to a better comprehension of the cognitive mechanisms involved when humans interact with the physical reality they inhabit.

Understanding the cognitive models people use in reasoning about the physical world is an important issue for Cognitive Science. Moreover, mental models theory provides many attractive features for developing computational models and systems able to manipulate physical situations (Geminiani et al., 1996; Forbus and Gentner, 1997):
(i) Mental models are dynamic structures which are created on the spot to meet the demands of specific problem-solving situations.
(ii) They perform the task of exploring complex and unknown situations, in that they include structures and processes able to simulate the transitions that can occur in phenomena being modelled.

(iii) When the available knowledge is incomplete, they make assumptions by means of a revision process in order to complete the explanation, each assumption leading to the construction of a mental model containing a hypothesis about the causes and dynamics of the physical situation.

By means of these features, it is possible to predict the onset and course of physical processes, to find out how to influence, control and prevent them, to diagnose unusual events, to allow the exploration of complex and unpredictable aspects of the physical phenomenon, and to reveal the incorrectness and/or incompleteness of the background knowledge. For these reasons, the physical process manipulation kernel of the illustrated computational system has been tested as an automated support for psychological experimentation, both within a toxicology framework (Carassa et al., 1995) and as a control system for managing urban traffic situations (Bandini et al., 1997a; 1997b).

2.1 Mental Models and Causal Models

Experimental psychology has demonstrated that humans and most superior mammals (Sperber et al., 1995) perceive causality when they are observing objects that come into direct contact. The term *perception* denotes that no cognitive process is requested; in fact, infants, 4 and a half months old have the same interpretation of causal sequences as adults: this kind of perception is automatic, obliged and innate (Leslie and Keeble, 1987; Michotte, 1946).

Geminiani et al. (1996) maintain that contact between objects is the crucial aspect of physical causality, not only in perception, but also at the cognitive level of causal analysis. The basic idea is that when human subjects are reasoning about two events which they judge to be causally linked, they try to represent a physical interaction between objects. There is much concern about assigning causal roles to two specific objects (AGENT and TARGET) and envisaging how they come into contact. *Contact* is seen as the necessary condition for the realisation of the causal link between events at a more abstract level.

In order to reason about the dynamics of a physical process, it is necessary to postulate the presence of a MEDIUM, an object or a set of objects that allows an AGENT to act on a TARGET in spite of their spatial non-contiguity. The causal process is the process that occurs first and allows contact between AGENT and TARGET be performed. A set of MODIFIERS, i.e. elements able to influence causal reasoning, may also be considered.

According to Johnson-Laird (1983), there are six major types of mental models of the physical world, one of which is the "kinematic model", representing changes and motions of depicted entities. Dynamic models are a subset of kinematic models and represent causal relations between certain events in a temporal sequence. Geminiani et al. (1996) provide a model to explain causality by contact, concerned only with a subset of dynamic models *–causal models–* in which causal events are represented as physical objects and processes.

In the following, we will adopt a very naive but intuitive example as a case of causality by contact (Carassa et al., 1995): the poisoning of a person caused by a snakebite. The poison (AGENT) is represented as an aggregate of particles carried by the blood flow within the circulatory system (MEDIUM) towards the heart (TARGET). MODIFIERS are elements influencing the path of the poison from the bite to its final destination.

Causal models include:

- a *structural component* that represents objects, their physical behaviour and physical relations;
- a *modifiable component* that represents qualitative parameters of objects, their behaviour and their relations.

Qualitative parameters include:

- physical properties (e.g., the position of the bite and size of the vein);
- spatial and temporal relations (e.g., the distance between the heart and the point of entry of the poison or the time between the snakebite and the injection of an antidote).

Specifying the qualitative parameters corresponds to the construction of a specific model. A reasoning process starts with the construction of a BASE MODEL, the simplest model representing the physical process without MODIFIERS, and with standard values of the qualitative parameters, taken as a reference system for all other models. Alternative models can be constructed by varying these parameters and by introducing MODIFIERS.

It is important to note that a causal model is a dynamic model that develops with time, simulating how involved objects interact.

2.2 Reasoning processes

The starting point of reasoning about physical phenomena is to assign causal roles to two specific entities and simulate how they can come into contact. Considering how the contact can be realised, as well as facilitated or hindered, is essential for planning how to modify or direct the course of the events. Mental models are a representational structure able to support exploration and explanation of physical situations as reported above.

In accordance with the spirit of the mental models theory, the inferential processes involved in common sense reasoning are the following:

- a ***construction*** process: takes as input premises expressing causal events, generating mental models of involved physical entities and processes;
- a ***matching*** process: has dynamic models as input, carrying out a series of comparisons between them and producing as output a first conclusion;
- a ***revision*** process: evaluates the extent to which the first conclusion can be considered acceptable, by searching for alternative models.

Though based on mental models, reasoning processes always rely on examples, a mental model being a representation of a specific situation. To verify the general validity of a reasoning process based on the analysis of particular cases, it is necessary to generate a finite set of significant examples using "falsify" procedures on mental models (Johnson-Laird, 1983).

3. Causal Models and QPT

Providing a computational representation of causal models is crucial, in that the availability of a formal description language is the starting-point for developing a computational system for reasoning about physical situations. We adopted a logical interpretation of Qualitative Process Theory -QPT- (Forbus, 1984), proposed by Bandini et al. (1988), in order to capture in a computer program the main tenets of the mental models. In fact, QPT preserves their main characteristics and, moreover, it allows for passing from an intuitive description of mental models to a fully worked-out computational model. QPT has many advantages for representing the mental models of the physical world in computational terms.

- It allows various common sense physical domains to be represented and treated.
- It supports the qualitative approach, typical of common sense reasoning, to represent parameters characterising entities involved in mental models; moreover, it provides an accurate representation of the qualitative parameters of individual entities and of the relationships between them.
- A causal model is a dynamic model that develops with time, simulating how involved entities interact. QPT allows the dynamic aspects of objects to be represented by means of the notion of *physical processes*.
- It allows causal relations between objects (AGENT, MEDIUM, TARGET and MODIFIERS) to be explicitly represented by means of qualitative *dependencies*. For a discussion about explicit and implicit descriptions of causality between qualitative parameters, see Forbus and Gentner (1986).
- It introduces explicit representation and treatment of *actions*, which allow new entities and processes to be introduced during the simulation of a physical model. This supports the construction of alternative models during the revision process.

3.1 Representing causal models

Now we present a Qualitative Process Theory (QPT) computational model based on a *language -QPL-*, for the description of physical domains and situations, and an *interpreter* for QPL *-IQPL-*, providing the main primitives for reasoning about the descriptions of physical systems and domains. The interpreter is implemented in Prolog language.

Qualitative Process Theory is a formal instrument for representing knowledge and supporting common sense reasoning about the dynamics of physical situations. A qualitative description of a physical situation implicitly embeds common sense knowledge on the behaviour of the situation itself, namely its dynamic description. Two elements are needed for the inferential apparatus of QPT to render this knowledge explicit.

- *A knowledge base*: common sense knowledge on the domain to which the situation belongs. It comprehends individual objects involved, their properties and the relations between them. Moreover, the description of the processes that can occur in such a domain is essential. In QPT, it is a non-changeable type of knowledge. In our system, it can be revised and augmented by means of a constructive induction process.

- *A scenario*: the description of a physical situation belonging to the domain at a given time. The scenario is a set of logical predicates, which are considered to be simultaneously true.

QPLanguage	**Mental Model**
entity *individual view*	objects and their physical properties
process *action*	behaviour and interaction between objects
scenario	initial state of the model

FIGURE 1: Correspondence between mental models of the physical world and QPL

The fundamental feature which marks this reinterpretation of QPT that we used is its logical conception: QPL, in fact, is a language entirely based on predicate logic. The figure above shows the correspondence we adopted between QPL and causal models.

Following is a brief description of main components of QPL models. *Entities*. These represent the individual entities of causal models. In a poisoning situation, entities are the poison, the antidote and so on. Relevant characteristics of entities are represented by qualitative variables, which take on values in the totally ordered set of symbols:

```
{infneg, highneg,..., zero, lowpos,..., infpos}
```

This set is called the *Qualitative Quantity Space(QS)* (De Kleer and Brown, 1984), where every qualitative value corresponds to an interval of real numbers, except for zero, which corresponds to a number. The granularity of the QS depends on the considered physical domain. *Processes*. The behaviour of entities can be represented by processes; examples are the flow of poison within the circulatory system and the action of some MODIFIERS on this flow. Processes are mechanisms that act on entities by changing their qualitative parameters. *Influence* is the main component of a process and represents direct causes of change. Indirect causes are represented by means of dependencies (see below). QPL represents influences using the syntax:

```
influence(increase,Intensity,Parameter)-positive
influence-
influence(decrease,Intensity,Parameter)-negative
influence-
```

The value of a changing Parameter increases when influenced by a positive influence and decreases when affected by a negative influence.
Dependencies. These express indirect causes of change. If a process, by means of an influence, *directly* affects some parameter R and some other parameter Q is qualitatively proportional to R, then we say that R *indirectly influences* Q. In QPL they are represented by the syntax:

```
dependency(Direction, Type, Quantity1, Quantity2)
```

where `Direction` indicates direct or inverse direction of the dependency, and Type indicates the type of dependency between `Quantity1` and `Quantity2`. Type takes values within the totally ordered set of symbols:

```
{log, rad, lin, poli, exp}
```

Actions. These describe instantaneous changes which can occur during a process and their structure derives from STRIPS (Fikes et al., 1971). This distinctive characteristic of QPL (Bandini et al., 1988; Forbus, 1989) introduces the possibility of representing and handling changes within the execution of a process. For example, it allows the introduction, at various times during the simulation of a process, of new entities, which are not present when the process starts. Effects of an action persist until simulation ends or a new action is introduced, which restores the previous state.

With respect to the cognitive model, TARGET, AGENT and MEDIUM are represented in terms of QPL entities. MODIFIERS can be represented both by entities interacting with AGENT and by processes and actions.

In the following a poisoning model in QP language is presented.

```
poison(Name) is_a entity with
   quantities: poison_in_vein(Name)
   and relations:
      poison_in_vein(Name)
      greater_than_or_equal_to_zero.

vein(Name, E) is_a entity with
   quantities: length(Name) & width(Name)
   and relations:
      dependency(inv,poli,width(Name),blood_speed(E)).

blood_flow(Name, E) is_a process with
   individuals: poison(Name) & vein(Name, E)
   and preconditions: alive
   and influences:
      influence(increase,tot(E),poison_in_heart(Name))
      &
      influence(decrease,tot(E),poison_in_vein(Name)).

tourniquet(T, Name, E) is_a action with
   individuals: vein(Name, E)
   and preconditions: alive & not_tourniquet
   and add_list: yes_tourniquet &
      dependency(linear,tot(E),tightness_degree(_,X))
      & influence(decrease, tot(T),width(Name))
   and delete_list: decrease(tot(T), width(Name)).

snakebite(Name, E, O) is_a scenario with
   individuals: poison(Name)   &   vein(Name,    E)    &
   heart(E)
   and initial_values:
      alive & poison_in_heart(Name) is_equal_to zero &
      not_tourniquet & not_antidote & not_cardiotonic
   and end_conditions:
      dead ¦ poison_in_vein is_equal_to zero.
```

3.2 Simulating causal models

A program devoted to qualitative reasoning on a represented physical situation can be intuitively conceived of as an "interpreter" of such a representation, on which it must be able to operate suitable manipulations. A QPT interpreter manages the dynamics of the physical situations described in a model, by simulating the behaviour of the entities involved and their interaction. Behaviour and interaction are represented as processes that act on entities and between them by modifying their qualitative parameters.

In order to simulate models of physical situations, the IQPL –Interpreter of Qualitative Process Language– follows two steps.

1. *Activation of starting conditions* - It activates entities, processes and actions whose conditions are verified (or deactivates them).

2. *Modification of qualitative parameters* - Until a steady state is reached (a steady state corresponds to the exhaustion of particles of poison or to the reaching of the deadly threshold), IQPL:

- changes parameters of entities which thus evolve from the current to the next state;
- resolves conflicting influences acting on a single entity and determines the resulting direction of change;
- propagates changes in a parameter to all other parameters functionally dependent on it.

4. Architecture of the System

According to Johnson-Laird (1993), induction occurs in three stages. The first stage is to grasp some propositions, some verbal assertions or perceptual observations. The second stage is to yield a mental model (hypothesis) that reaches a better description or understanding of this information in relation to a background of general knowledge. Such models have to be consistent with both such observations and background knowledge. The third stage is to evaluate the hypothesis and as a result, to maintain, modify or abandon it.

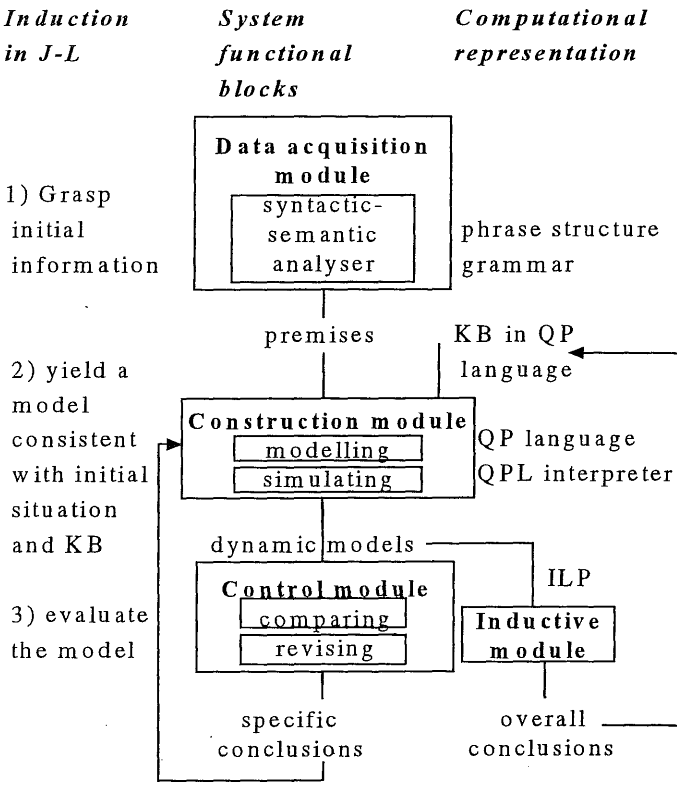

FIGURE 2: System description

The overall system architecture reflects the proposed phases. Their modules and their correspondence to the main phases are depicted in Figure 2.

4.1 Syntactic-Semantic Analyser

This allows introduction of natural language sentences, which give information about physical situations that can occur within a particular physical domain. It is based on a *phrase-structure grammar* proposed by Chomsky (1957) (for a detailed description, see Carassa et al., 1995). As output, the analyser carries out the scenario, providing the construction module with the description of the physical situation that has to be analysed.

4.2 Construction Module

In order to draw inferences about physical situations, the system builds models of them, making use of data drawn from:
(i) A *scenario*, representing the initial information provided to the system and describing a physical situation belonging to the domain at a given time.
(ii) The *knowledge base*, expressing the common sense knowledge on the domain to which the situation belongs, represented in QP Language.

Modelling procedures, on the basis of the initial information, build QPT models of described situations by activating entities and processes from previous knowledge. Moreover, they assign values to each of the qualitative parameters characterising entities, their relationships and processes.

Simulating procedures, by means of the Interpreter QPL, represent the dynamic evolution of the physical situation described in QPT models, by simulating the entities' behaviour and the interaction between them.

Every simulation results in a final state of the specific situation. In a causal model, the final state regards the fact that contact between AGENT and TARGET could take place or not and, in case of contact, the qualitative characteristics of the contact itself are specified. In the case of poisoning, the main parameter characterising the final state is the amount of poison in the heart: this amount may be lower or higher than the lethal threshold.

A previous implementation of the CONTROL and CONSTRUCTION MODULE has been presented in Carassa et al. (1995).

4.3 Control Module

The system draws backward inferences: when in a problem-solving context it recognises the initial and the final state of a situation, it tries to reconstruct the intermediate events. In the case of poisoning, when the initial state –snake-bite– and the final –death or survival– are both given, the aim is to define some enabling conditions –i.e. the injection of an antidote– to evolve the model in the desired way.

As input, the CONTROL MODULE receives the causal model and its evolution from the CONSTRUCTION MODULE. If the correspondence between the final state of the model produced by the CONSTRUC-

TION MODULE and the conclusion prescribed by the premises is not fulfilled (see example below), the CONTROL MODULE searches the various enabling conditions that would allow the model to approach the fixed final state. It follows that in order to draw a backward inference, the system generates different forward simulations, until the desired final state is reached; in fact, given an initial and a final state of a physical situation, there are different ways of connecting them. For this reason, the CONTROL MODULE introduces modifications to the model. In this way, new models are recursively built until their final state is compatible with the final state defined by the initial premises. The CONTROL MODULE is composed of *matching procedures* and *revision procedures*. The last state how alternative models can be generated. Two strategies are applied:

- new entities (MODIFIERS) are introduced, which can progressively be added to the initial model, so as to originate increasingly complex models;
- qualitative parameters of the already represented entities are systematically changed.

It is important to stress that, when the correspondence between the causal model and the initial premises is not fulfilled, there is no explicit rule that prescribes how a model should be modified so as to approach the fixed final state. Revision procedures proceed by trial and error. They introduce in succession only one change. Pairs of models, which differ one from the other in a single characteristic, are matched by matching procedures, in order to discover the precise effect of the change; in fact, the matching procedures find out how two dynamic models differ in their qualitative parameters and in their final states. Revision proceeds by the selection of changes that lead the final state of a model closer to the desired situation.

From a computational point of view, QPT allows new entities (MODIFIERS) to be introduced during the revision of a causal model, by means of explicit *actions* activated during qualitative simulation of processes.

Example of model generation

Now we describe an example of backward inference. A situation is presented where Cleopatra survives a snake-byte and has been applied a not very tight tourniquet. The goal of the system is to find out the enabling conditions justifying such an evolution. With this aim, the system builds, revises and matches a succession of models, until a model is found whose conclusion coincides with the prescribed final state.

Premises:
Cleopatra was bitten by a very poisonous snake;
Cleopatra was applied a not very tight tourniquet;
Cleopatra survived.
Model 1
Conclusion: the amount of poison in the heart is greater than the lethal threshold. Cleopatra died.
$>>>$Premises are not verified$<<<$

Since the final state of MODEL1 does not correspond to the result mentioned by the premises, the system applies revision procedures that lead to the construction of an alternative model –MODEL2–, where the tourniquet is applied after a long period of time.

Model 2
Revision by trial: the tourniquet was applied after a long time from the bite
Conclusion: the amount of poison in the heart increases. Cleopatra died.
$>>>$Premises are not verified$<<<$; goal is further away

Since the final state of MODEL2 is further than the one of MODEL1 from the prescribed final state, MODEL2 is rejected. Comparing MODEL1 and MODEL2 guides the system into the next revision.

Model 3
Revision guided by comparing Model1 and Model2: the tourniquet was applied a short time after the bite
Conclusion: the amount of poison in the heart decreases. Cleopatra died.
$>>>$Premises are not verified$<<<$; goal is closer

Since MODEL3 is nearer to the goal, the modifying path of the revision is correct. Following it, MODEL4 reaches the prescribed conclusion.

Model 4
Revision guided by comparing Model2 and Model3: the tourniquet was applied

a very short time after the bite
Conclusion: the amount of poison in the heart is lower than the lethal threshold. Cleopatra survived.
>>>Premises are verified<<<

The repeated revision of the models and the evaluation in relation to the premises allow the system to construct a model which is consistent with both the background knowledge represented in the system and the information provided by the premises. This model represents a plausible explanatory hypothesis about the causes and the dynamics of the physical process generating the situation. Such a hypothesis is specific in that it is valid only for the particular situation described by the initial premises.

4.4 Inductive module

In order to yield overall hypotheses covering classes of situations, the system analyses the series of successive models of the physical domain generated by the revision process.

For the evaluation of hypotheses and the enrichment of background knowledge, we turned to recent advancements in the Machine Learning field, whose methods are natural candidates for providing computational tools able to model human learning. Specifically, our approach being based on a logic representation, we made use of results from Inductive Logic Programming (ILP - Muggleton, 1991; Muggleton and DeRaedt, 1994), a research area rooted in computational logic and inductive machine learning. Prominent characteristics of ILP are: (i) the ability to induce hypotheses from observations (examples); (ii) the use of substantial background knowledge in the learning process and the use of such knowledge as essential for guiding the computational system to achieve effective domain knowledge.

With this aim, the series of tentative explanatory models is first considered as a set of examples. A mental model can be viewed as an example, in that it always represents a specific situation: it is not a general structure, but is constructed for the particular goal pursued.

In particular, we state the comparative analysis of models in terms of a constructive induction problem (Idestam-Almquist, 1992). This problem is formally defined in the following way. Given:
(i) Knowledge Rules (KR)

(ii) Case Facts (CF) and
(iii) a Reached Conclusion (RC),
such that: KR $\cup$ CF $\nvdash$ RC, find IR (Inductive Rules) such that: KR $\cup$ CF $\cup$ IR $\vdash$ RC.

Informally, this means that, whenever the background knowledge is insufficient to explain (in logical terms, to derive) the conclusion observed in the case at hand, additional rules are inductively searched for, which, when added to the background knowledge, allow the derivation of the conclusion.

In our framework for mental models of physical situations, we have the following correspondence with the terms of the constructive induction problem: KR corresponds to the BASE MODEL (§ 2.1); CF corresponds to the particular facts described by the premises, characterising the initial state of the physical situation; IR are the overall hypotheses expressed as general rules; RC corresponds to the final state prescribed by the premises.

Thus we have: given a set of examples (models) and a background knowledge (domain knowledge of the system), the goal is to find a hypothesis that explains the examples with respect to the background knowledge. Discovering such hypotheses can lead to an enrichment of the knowledge domain, in that they reveal new causal relations between existing entities in a mental model. By means of such causal relations, the system finds out how to influence, initiate or prevent the physical phenomenon represented in the models and how to diagnose unusual events.

The simulation process constructs models by finding, for each set of parameter values, specific enabling conditions leading from the initial state to the final state of that model. Using these results, the constructive induction process constructs new rules expressing general enabling conditions over the set of all models. The overall hypotheses are therefore expressed as general rules covering classes of situations and the background knowledge of the system is subject to these new rules. Consequently, the system may flesh out a model of a new situation with the additional information that is automatically provided by its knowledge, without regenerating a series of tentative models.

An inductive logic program has been used, which was previously developed (Baroni, 1997) and experimented in another domain (Costantini and Lanzarone, 1996).

In the case of the analysis of a MODIFIER, the system can find out how, by varying its parameters, to influence the given situation. As an example of MODIFIER, a tourniquet is characterised by three parameters: degree of tightness, time elapsed before its application and point of application. A tourniquet is proximal, if it is placed between the snakebite and the heart, or distal, if placed before the bite.

In order to find out how a tourniquet has to be applied, the system systematically changes its parameters and observes the resulting effects on the evolution of the situation. The introduction of any change corresponds to the generation of a new model. The CONTROL MODULE retains such a succession of models, collecting for any of them the relevant parameters characterising its initial –corresponding to CF- and final state –corresponding to RC–. If in a model$_i$ a not tight tourniquet is applied in proximal position a very long time after the bite, a great amount of poison accumulates in the heart. The initial and final state in such a model are the following:

Initial state (Case Facts)
tightness_degree(model$_0$, lowpos).
time_application(model$_0$, highpos).
point_of_application(model$_0$, proximal).

Final state (Reached Conclusion)
poison_amount(model$_0$, highpos).

During the revision phase and the exploration of the effects of a tourniquet, a set of models are collected:

Initial states (Case Facts)
tightness_degree(model$_1$, lowpos).
tightness_degree(model$_2$, medpos).
tightness_degree(model$_3$, highpos).
. . .
time_application(model$_1$, highpos).
time_application(model$_2$, lowpos).
time_application(model$_3$, lowpos).
. . .
point_of_application(model$_1$, proximal).
point_of_application(model$_2$, proximal).
point_of_application(model$_3$, distal).
. . .

Final states (Reached Conclusion)
poison_amount(model$_1$, highpos).
poison_amount(model$_2$, lowpos).
poison_amount(model$_3$, highpos).
. . .

The ILP program comparatively analyses this set, in order to induce new relationships between the parameters of the tourniquet and the amount of poison in the heart. In this way, it can discover under which conditions a tourniquet effectively bars a poisoning process. The amount of poison in the heart is the lowest when a very tight tourniquet is applied after a very short time in proximal position. A not very tight tourniquet, however, can be effective if it is applied after a very short time, as is shown in the previous example (§ 4.3).

The new relationships are expressed as general rules (containing variables instead of constants only). For example, the following rule states that a tourniquet applied in distal position does not act against the poisoning process – there is a great amount of poison in the heart – whatever the values of the other parameters may be (Model is a variable, thus covering a class of specific models):

poison_amount(Model, tourniquet, highpos):- *(Inductive Rule)*
    application_point(Model, distal).

The base knowledge of the system is subject to the new rules. In this knowledge, the tourniquet is represented as a QPT action:

```
tourniquet(T, Name, E) is_a action with
    individuals:   person(cleopatra) & vein(Name, E)
    and preconditions: alive & not_tourniquet
    and add_list:
        dependency(linear, tot(E),tightness_degree(_,X)) &
        influence(decrease, tot(T), widht(Name)) &
        yes_tourniquet
    and delete_list: not_tourniquet.
```

During the simulation phase, the QPT interpreter carries out the instructions reported in add_list and delete_list, if and only if the individuals and preconditions are verified. This corresponds to the fact that the tourniquet decreases the width of the vein proportionally to its degree of tightness, when Cleopatra is alive and she has not yet been applied a tourniquet. For example, the discovery of the inductive rule above enriches the knowledge base about the application of a tourniquet; the new information in inserted as a precondition:

```
tourniquet(T, Name, E) is_a action with
    and preconditions: alive & not_tourniquet &
        point_of_application(_, proximal).
```

In this way, during the analysis of new physical situations, the system will immediately apply the tourniquet in a correct position without having to generate several models in order to find out effective points of application.

5. Concluding remarks

In conclusion, we take the position that the representation of the mental models theory within a computational framework - specifically in the context of qualitative reasoning about physical systems - leads to refinements of the general theory and to a more precise distinction between different aspects of a learning process. In fact, what is generally called induction in Johnson-Laird has been differentiated into two phases. The first one is the generation by simulation of incremental models and the second one is the achievement of a revised and augmented knowledge by means of a constructive induction process applied to generated models. We believe that the identification of the distinction between these phases and their interaction represent a new point of view in studying model revision, inductive analysis and learning in a cognitive perspective.

As a possible future development of this line of work, we will investigate the recognition of new causal relations emerging from the induced knowledge. Additional further developments will single out some real life application domains, in order to validate the proposed approach.

University of Milan
Insubria University (Varese)
University of Turin

REFERENCES

Bandini S., Bruschi M., Filippini M.G., Molesini A. (1988), An interpretation of QPT computational model according to logic programming, in: *Computational Intelligence*, A. Martelli, G. Valle, eds., North Holland, Amsterdam.

Bandini S., Carassa A., Pegurri P., Valpiani A. (1997), Modeling and simulating traffic from a cognitive and qualitative reasoning perspective. *Proceedings of the 11th International Workshop on Qualitative Reasoning,* Cortona, Italy.

Bandini S., Carassa A., Pegurri P., Valpiani A. (1997), Traffic simulation based on a cognitive and qualitative reasoning model. *Proceedings of the 3rd International Joint Conference on Artificial Intelligence Workshop on Engineering Problems for Qualitative Reasoning,* Nagoya, Japan.

Baroni A. (1997), An ILP system in Reflective Prolog for learning legal rules, *Degree Thesis, Computer science Dept.,* University of Milan (G.A. Lanzarone and S. Costantini supervisors) (in italian).

Carassa A., Valpiani A., Geminiani G., Bandini S. (1995), A cognitive model of causal reasoning about the physical world. In: *Topics in Artificial Intelligence*, M. Gori, G. Soda, eds, Springer Verlag, Lecture Notes in Artificial Intelligence 992, Berlin.

Chomsky N. (1957), *Syntactic structures*, Mouton, The Hague.

Costantini S., Lanzarone G.A. (1996), Statutory Interpretation and Legal Analysis using Explanation-Based Generalization and Constructive Induction, *Proceedings of the 5th Iberoamerican Congress on Informatics and Law,* La Habana,Cuba.

D'Ambrosio B. (1989), Extending the Mathematics in QPT, *International Journal of Intelligent Systems* 4: 55-80.

de Kleer J., Brown J.S. (1984), A Qualitative Physics Based on Confluences, *Artificial Intelligence 24,*7-83.

Di Sessa A.A. (1982), Unlearning Aristotelean Physics: a Study in Knowledge-Based Learning, *Cognitive Science* 6.

Fikes R.E., Nilsson N.J. (1971), STRIPS: a new approach to the application of theorem proving in problem solving, *Artificial Intelligence 2:* 189-208.

Forbus K.D. (1984), Qualitative Process Theory. *Artificial Intelligence* 24: 85-168.

Forbus K.D. (1989), Introducing Actions into Qualitative Simulation, *Proceedings of the 11th International Joint Conference on Artificial Intelligence,* Detroit, MI.

Forbus K.D., Gentner D. (1986), Causal Reasoning about Quantities, *Proceedings of the 5th Annual Conference of the Cognitive Science Society,* Amherst, MA.

Forbus K.D., Gentner, D. (1997), Qualitative Mental Models: Simulations or Memories? *Proceedings of the 11th International Workshop on Qualitative Reasoning,* Cortona, Italy.

Geminiani G., Carassa A., Bara B.G. (1996), Causality by contact, in: *Mental Models in Cognitive Science*, D. Garnham, J. Oakhill, eds., Lawrence Erlbaum, Hillsdale, NJ.

Gentner D., Stevens A., eds. (1983), *Mental Models.* Lawrence Erlbaum Associates Publishers, Hillsdale, NJ.

Hayes P.J. (1979), The Naive Physics Manifesto, in: *Expert Systems in the Microelectronic Age*, D. Michie, ed., Edinburgh University Press, Edinburgh.

Holland J.H., Holyak K.J., Nisbett R.E., Thagard P.R. (1986), *Induction: Processes of Inference, Learning and Discovery,* MIT Press, Cambridge, MA.

Idestam-Almquist P. (1992), Learning Missing Clauses by Inverse Resolution, *Proceedings of the International Conference on 5th Generation Computer Systems,* Tokio, Japan.

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

Johnson-Laird P.N. (1993), *Human and Machine Thinking*, Lawrence Erlbaum, Hillsdale, NJ.

Leslie A.M., Keeble S. (1987), Do six-month-old infants perceive causality?, *Cognition* 25: 265-288.

Michotte A. (1946), La perception de la causalite', Louvain, Ed. de l'Institut Superieur de Philosophie (Etudes de Psychologie), VIII. Trad. (1963). *The perception of causality*, Methuen, London.

Muggleton S. (1991), Inductive Logic Programming, *New Generation Computing* 8 (4): 295-318.

Muggleton S., DeRaedt L. (1994), Inductive Logic Programming: Theory and Methods, *Journal of Logic Programming* 19/20: 629-679.

Sperber D., Premack D., Premack A.J., eds. (1995), *Causal Cognition. A Multidisciplinary Debate*, Clarendon Press, Oxford.

Weld D., De Kleer J., eds. (1990), *Readings in Qualitative Reasoning About Physical Systems*, Morgan Kaufmann Publishers, San Mateo, CA.