Conceptual Change in the Explanations of Phenomena in Astronomy

Katharina Morik, Univ. Dortmund
Martin Mühlenbrock, Univ. Duisburg

1 Introduction

Childrens' explanations for astronomical phenomena have been well investigated for years (e.g., [Baxter, 1989],[Nussbaum, 1989a]). A detailed developmental study of children's explanations of the day/night cycle has been conducted by Vosniadou and Brewer ([Vosniadou and Brewer, 1992],[Vosniadou and Brewer, 1993], [Vosniadou and Brewer, 1994]). The collected data of this study together with their interpretation from the point of view of developmental psychology offer an excellent basis for the investigation of conceptual change. The day/night cycle involves several objects (the earth, the sun, the moon, the observer) and is characterised by their spatial relations that change over time. Children's explanations of the day/night cycle reveal that they are embedded within an naive theory about the physical world, which is based on interpretations of everyday experience. The way from naive explanations, directly related to observations, to the scientific explanation involves dramatic revisions: the idea of the earth as an infinite ground has to be changed to the idea of the planet earth. The complex learning task has to be accomplished on the basis of instruction alone, as there is no everyday experience or task that is in conflict with the naive explanation. The difficulty of this learning task can be characterized with respect to four dimensions. First, it deals with theories that relate several concepts. This corresponds to the view of Murphy and Medin [Murphy and Medin, 1985]. Second, it covers several types of learning. In addition to concept acquisition, conceptual change is studied, where both are integrated into problem solving (here: explaining the day/night cycle). Third, the theory to be developed is not isolated or artificial, but concerns everyday experience. Fourth, in addition to observations of day and night, linguistic input including erranious explanations underly theory development. Finally, it should be noted that this complex learning activity takes years and is best considered as life-long learning.

Given the characteristics of learning astronomical phenomena, it is obvious, that the learning process can hardly be investigated considering explanations of one child at one point in time under controled laboratory conditions. Is is also clear, that an accompanying study of one child over years is impossible. Hence, the only empirical basis for any investigation are inquiries of children of different age. The interviews are snap shots of an ongoing learning process. For the scientific investigation of life-long learning, the snap shots form landmarks. In order to explore the paths between the landmarks, we have started an interdisciplinary investigation [Morik, 1996].

A computational model of the observed knowledge states has been built. This includes the formal modelling of spatio-temporal concepts and of using the concepts for problem solving (here: forming an explanation for observations on demand). The formal model makes explicit an understanding of the empirical data (the children's explanations). It raises questions and points at missing data or knowledge about what the children actually had in mind when giving the recorded answers in the inquiry. The computational model must reproduce the observations of the empirical study, i.e. it has to meet the landmarks. Then, learning operators in the sense of Siegler [Siegler, 1991] had to be formalized. Now, systematic experiments could be undertaken that are impossible with human subjects. Different learning operators and different inputs into the simulated theory could be applied in different orders. Since the topic of the interdisciplinary investigation are the cognitive processes that lead from one knowledge state to the next (the paths between the landmarks), all transitions between the formally modeled knowledge states have been simulated. It is then interesting to characterize those transitions that correspond to psychologically assumed actual transitions. What distinguishes these from other transitions that seem to not happen in children's thinking? Are there formal distinctions that clearly seperate simulated transitions which correspond to children's development from those which do not correspond to children's development? Can these formal characterizations be interpreted in psychological terms?

Note, that the computational model does not play the role of an artificial child that shows all aspects of a child's behavior! Instead, the computational model plays the role of a description of children's thinking in almost the same way as does a textual psychological description. In much the same way as is a psychological article, a computational model is focused on a particular question. If the development of astronomical theories in children is the research topic, neither the article nor the computational model cover how children learn to grasp, recognize objects, or creep. Both, the author of the article and the author of the computational model, of course, agree in the importance of such learning steps for children [1]. It is just, that the current investigation is about theory change. Hence, the computational model of explanations of the day/night cycle and sequences of transitions between different explanations does not model phenomena which are not covered by the inquiry of children.

When starting the interdisciplinary work, our hypothesis was that the transition between different explanations for the day/night cycle would require knowledge revision and a change of the representation language. However, we found out that most of the knowledge remains the same when transiting from one explanation to the next, deductive inference and knowledge revision dominates the evolution of the scientific explanation, no representation adjustment is necessary. In this chapter, we explain these surprising findings. Moreover, we give a formal account for the difficulty of the conceptual changes observed in the empirical study and simulated in the computer model. The remaining of the chapter is divided into four sections. First, the empirical study is presented. Second, the representation formalism and the inferential operators of the system we used for the computer simulation are described. Given these prerequisites from the two disciplines psychology and artificial intelligence, we describe in the fourth section the computational model, its validation and experiments. We discuss possible generalizations of our case study and point at open research questions in the last section.

[1] Learning the mapping from perceptions to linguistic concepts has been computationally investigated using a mobile robot and the same learning system as has been used in the day/night cycle project [Klingspor et al., 1997],[Rieger, 1996].

2 An empirical investigation of the development of childrens' explanations of the day/night cycle

In recent years, a number of psychological studies investigated individual conceptions of phenomena in observational astronomy. These studies investigated children's conceptions of celestial bodies such as the earth, the sun, and the moon and their conceptions of processes such as the alternation of day and night and the waxing and waning of the moon [Klein, 1982, Viglietta, 1986, Sadler, 1987, Baxter, 1989, Vosniadou and Brewer, 1994]. Not only is it a relatively rich knowledge domain composed of a number of concepts with more or less complex relations, but also does children's everyday experience provide them with sufficient information to develop an intuitive understanding of many of the phenomena that are part of the domain of scientific astronomy. Furthermore, children get in contact with scientific information on this subject from early on, as it is part of primary schools curricula. However, scientific conceptions of astronomical phenomena differ from children's intuitive conceptions bnased on perceptions in some major respects. This makes the influence of scientific information on those intuitive conceptions a center of interest.

The subjects in the psychological study were 60 children from different grades (mean ages 6.9 in first grade, 9.9 in third grade, and 11.0 in fifth grade), attending an elementary school in Urbana, Illinois. By means of a 48-item questionnaire based on extensive pilot work, the children were individually interviewed on their concepts in astronomy. From the overall 48 questions, only 13 questions investigated children's explanations for the day/night cycle, including their ideas about

1. the sun (e.g. Q 22: "Where is the sun at night?", Q 23: "How does this happen?"),
2. the alternation of day and night (e.g. Q 25: "Now make it so it is day for that person. Good! Now make it so it is night for that person."),
3. the movement of the moon (Q 30: "Does the moon move?", Q 33: "Why does the moon move?"),
4. and the disappearance of the stars during day (e.g. Q 36: "Where are the stars at night? Where are they during the day?"),

but excluding their ideas about the earth's shape and gravity. According to the four sets of questions, each child was classified by means of four seperate sets of categories by two independent judges, discussing all disagreements until consensus was achieved. For each child, the four different categories were combined afterwards and assigned to an overall mental model of the day/night cycle, independently from and without knowledge of children's mental models of the earth, which had been derived in a seperate evaluation.

3 The representation formalism and inferential operators – MOBAL

The simulation of the complex learning task as described above demands a powerful system that offers:

	Explanation Types	Class	Grade 1st	3rd	5th	Total	
1.	*The sun is occluded by clouds or darkness.*	initial	2	1	1	4	7 %
2.	*The sun and the moon move up/down on the ground.*	initial	7	0	0	7	12 %
3.	*The sun and the moon move to the other side of the earth.*	initial (synthetic)	2	0	0	2	3 %
4.	*Sun and moon move up/down unspecified as to earth side.*	initial	3	0	0	3	5 %
5.	*The sun moves out into space.*	initial	1	1	0	2	3 %
6.	*The sun and the moon revolve around the earth every day.*	synthetic	0	1	0	1	2 %
7.	*The earth and the moon revolve around the sun every day.*	synthetic	0	1	0	1	2 %
8.	*The earth rotates up/down; the sun and moon are fixed at opposite sides.*	synthetic	1	3	7	11	18 %
9.	*The earth rotates up/down; the sun is fixed but the moon moves.*	synthetic	0	1	3	4	7 %
10.	*The earth rotates around its axis; the sun and the moon are fixed at opposite sides.*	synthetic	0	1	1	2	3 %
11.	*The earth turns around its axis; the sun is fixed but the moon moves.*	scientific	0	1	0	1	2 %
12.	*The earth turns in unspecified direction; the sun is fixed but the moon may or may not move.*		1	1	1	3	5 %
13.	*Mixed: The earth rotates and the sun moves up/down.*		1	0	4	5	8 %
14.	*Mixed: The earth rotates and revolves.*		1	2	2	5	8 %
15.	*Mixed general.*		0	5	1	6	10 %
16.	*Undetermined.*		1	2	0	3	5 %
Total			20	20	20	60	100 %

Table 1: Children's explanations for the day/night cycle according to [Vosniadou and Brewer, 1994] (Explanantion type 3 could be classified as synthetic in having assimilated the scientific information on the existence of another earth side.)

- an inference engine that explains and predicts facts on the ground of given facts, so that the computational model can answer the same questions (translated into a formal language) that the children answered;
- induction operators that can interact with other operators, so that learning can be embedded in other processes of theory development;
- a procedure for automatically changing the representation, so that the transition from one explanation to the next can be modeled in terms of representation change;
- a knowledge revision that detects and resolves contradictions between (input and derived) facts, so that the input of a teacher can be recognized as a contradiction to the current explanation;
- a well-based formal semantics so that characterizations of the system's behavior correspond to formal characterizations.

MOBAL is such a system [Morik et al., 1993].

In this section, we introduce the terminology that we use when describing our computational model of explanations of the day/night cycle. Those who are familiar with the terminology of artificial intelligence in general and machine learning in particular will skim it in order to see how we have abbreviated the subtleties of a long scientific discourse on knowledge-based systems and machine learning. For those who are not familiar with artificial intelligence, we offer an informal introduction with pointers to standard logic. We hope that the informal explanations can intuitively be understood. The pointers to logic indicate a widely known formal basis. The general view is specialized by a short description of the MOBAL system.

3.1 Knowledge representation and inference

In order to create computer programs that learn, knowledge-based systems have been developed in artificial intelligence. These systems have a fixed set of operators and a fixed control structure. They change their behavior only because of changes in their knowledge-base. Given a number of facts and some inference mechanism, the rules derive new facts deductiveley from the given facts. See, for instance, the following knowledge base:

```
drink(ann,wdrink,event4) entails(wdrink,whiskey) entails(wdrink,water)
    contains(whiskey,alcohol)

drink(X,Y,E) & entails(Y,Z) & contains (Z,alcohol) → drunken(X)
```

This knowledge base can answer the question whether Ann is drunken:

```
?- drunken(ann)

yes
```

The main assumption of knowledge-based systems is: more knowledge (facts or rules) leads to better behavior. This assumption relates knowledge-based systems naturally with learning as the acquisition of more knowledge. Of course, there are other types of learning: increasing the efficiency of problem-solving, chunking, skill acquisition - to list just some of the types that have been discussed in literature. In this study, we focus on learning operators that induce more knowledge on the basis of given knowledge and on learning that takes the form of revision. There are diverse induction operators. They all model **conceptual change**. The major differences can be grouped into four aspects:

Supervised vs. unsupervised learning: if the input to learning consists of classified examples (comparable to the experiments of [Shepard et al., 1961]), it is called supervised learning. If categories are to be aggregated by the learning operator, it is called unsupervised learning.

Concept learning vs. rule discovery: if concepts, possibly a hierarchy of concepts, are learned in terms of their definitions, it is concept learning. If all valid and non-redundant rules are learned that underly a set of observations, it is called rule discovery.

Use of background knowledge: some learning operators are capable to enrich given examples or observations by given or already learned knowledge. This is more difficult than ignoring the background knowledge, since it asks for many deductive operations before induction starts. However, it can lead to better learning results.

An example introduced by Ryszard Michalski can illustrate this. Let the learning task be to find out what leads to a person being drunken. Given are classified instances of the concept of being drunken. In other words, the task is supervised concept learning. The instances given are:

```
drink(jim,gdrink,event1) entails(gdrink,gin) entails(gdrink,water)
    drunken(jim)
drink (bob,vdrink,event2) entails(vdrink,vodka) entails(vdrink,water)
    drunken(bob)
drink (ken,wdrink,event3) entails(wdrink,whiskey) entails(wdrink,water)
    drunken (ken)
```

A simple induction operator would learn the following rule:

```
drink(X,Y,E) & entails(Y,water) → drunken(X,E)
```

A more clever operator could use the following background knowledge:

```
contains(gin,alcohol) contains(vodka,alcohol) contains(whiskey,alcohol)
```

It could enrich the examples by these facts and learn the following rule:

```
drink(X,Y,E) & entails(Y,Z) & contains(Z,alcohol) → drunken(X)
```

We have seen above that this rule can be applied to new events whose description is not already classified by the predicate drunken.

Constructive induction: some learning operators are capable of adjusting the representation language. Since this issue is central to our topic, we elaborate on it.

First, we have to distinguish between representation formalism and representation language. The **representation formalism** determines the – possibly infinite – set of well-formed formulas. In the example above, the formalism is a restricted first-order logic. A term is either a constant term (written in small letters, e.g., jim) or a variable (starting with a capital letter, e.g., X). A fact is a positive or negated predicate whose arguments are terms. A rule is composed of premise and conclusion. The premise is a conjunction of facts. The conclusion is one fact. The deduction operators rely on this formalism. The **representation language** determines the signature (i.e. the set of predicates, in a many sorted logic with sort restrictions of the arguments) for a particular application of the formalism. In our example, the representation language is made of the predicates

```
drink(<Person>,<Drink>,<Event>) entails(<Drink>,<Ingredient>)
    contains(<Ingredient>,<Substance>) drunken(<Person>).
```

The argument sorts are written in brackets.

We can now explain constructive induction. The representation formalism remaining the same, constructive induction changes the representation language. There are several ways why and how to do so. In our example, we have not yet represented time explicitly. We forgot to add a time marking argument in the predicate drunken. This means that for all times the question about Ann being drunken will be answered positively. Constructive induction could introduce a new predicate new_p(<Person>,<Event>). If there were additional facts (e.g., about the dates of the events and their precedence relations), new rules like the following could be learned:

```
drink(X,Y,E) & entails(Y,Z) & contains (Z,alcohol) → new_p(X,E)

new_p(X,E1) & 12hours_later(E1,E2) → not(new_p(X,E2))
```

The newly introduced concept new_p(<Person>,<Event>) can then replace the inadequate concept drunken(<Person>). Constructive induction is a learning operator that automatically changes the representation language.

By **representation change** we mean the introduction of new predicates together with rules that relate the new predicate to other ones. Our definition of representation change covers the adjustment of a predicate, since we decompose this operation into the introduction of a new predicate and the replacement of the inadequate one [2]. Merely introducing new predicates into the representation language is in our view not sufficient for speaking of a representation change. We explicitly demand rules that – at least partially – define the meaning of the new predicate.

In addition to the deductive use of the knowledge-base for problem-solving and the induction of more knowledge, knowledge revision is an important operator. Several methods for knowledge revision exist and their theoretical properties have been investigated (e.g., [Gärdenfors, 1988]). The problem of knowledge revision is to keep a theory consistent when adding or deleting a statement of the theory or to make a currently inconsistent theory consistent by removing some statements. The theory is the inferential hull of a so-called base: the actual knowledge-base. Whereas most known revision operators with a well-defined formal

[2] The second step is actually not necessary. It may well be, that the inadequate concept remains in the knowledge base. There are even inadequate concepts that are nevertheless efficient for simple cases.

semantics and proven theoretical properties consider theories, a more recent approach allows to operate on explicit statements alone (base revision) and nevertheless keep the desired properties of the operator with respect to the resulting theory [Wrobel, 1994]. An example illustrates the main idea of Wrobel's base revision. Let us go back to our first attempt to model the concept of being drunken. If we want to add to the small knowledge base the fact that Ann is (no longer) drunken, we get at contradiction between the derived fact `drunken(ann)` and the input fact `not(drunken(ann))`. This contradiction is traced back to the rule and the instantiations of its premise on one hand and to the input facts on the other hand. Here, the derivation tree for `drunken(ann)` leads in one step to the rule and the input facts. It is then calculated, which minimal set of facts and/or rules could be removed in order to prevent the contradiction. The next step is the choice between the minimal removal sets. This can be performed by the user or by the system. The system uses the source of information: input fact, derived fact. If input facts are considered assumptions, the obvious revision is to delete one of them. If the facts are considered more reliable than derivations, then in this example only the rule can be removed. Since removing the rule is too far-reaching an attempt, a revised version of it is put back to the knowledge base.

```
not(instantiates(ann,X)) & drink(X,Y,E) & entails(Y,Z) & contains(Z,alcohol) →
    drunken(X)
```

This revised version of the rule excludes `ann` from the domain of `X` but is still applicable for all other instantiations of `X`. There could be other additional conditions which specialize the rule such that it no longer covers the negative input fact. For instance, additional facts about the quantity of alcohol could better characterize exceptions to the rule:

```
size(ann,large) quantity(wdrink,small) drink(ann,wdrink,event4) not(drunken(ann))

size(ken,small) drink(ken,wdrink,event3) drunken(ken)
```

If facts like this are known to the system, it can revise the rule further:

```
size(X,small) & drink(X,Y,E) & entails(Y,Z) & contains(Z,alcohol) → drunken(X)
```

Note, that because of the knowledge-base revision, the concept definition of being drunken has changed. In our small example, the revision was triggered by the predicate that represents the concept. It is possible, however, that the revision is triggered by one predicate and the revised rule defines another predicate, since the revision regards the derivation. Let, for instance, `drink(ann,wdrink,event4)` not be an input fact, but derived from the following rule

```
party(X,E) & offered(Y,E) → drink(X,Y,E)
```

and the facts

```
party(ann,event4)

offered(wdrink,event4)
```

The derivation of the contradicting fact `drunken(ann)` is now achieved using two rules and four input facts (see Figure 1). Using the assumption that input facts are more reliable than rules, each rule is a candidate for revision. The choice, which rule becomes revised depends on heuristic measures, e.g. how many successful rule applications are known. If the rule deriving `drink(X,Y,Z)` has not been successfully applied before, but the rule about `drunken(X)` has, the system will change the concept of `drink` becoming:

```
not(instantiates(ann,X)) & not(instantiates(wdrink,Y)) &
    not(instantiates(event4,E)) & party(X,E) & offered(Y,E) → drink(X,Y,E)
```

Indirectly, the conceptual change of drink also changes the concept of being drunken. Knowledge revision models conceptual change by taking into account the overall theory.

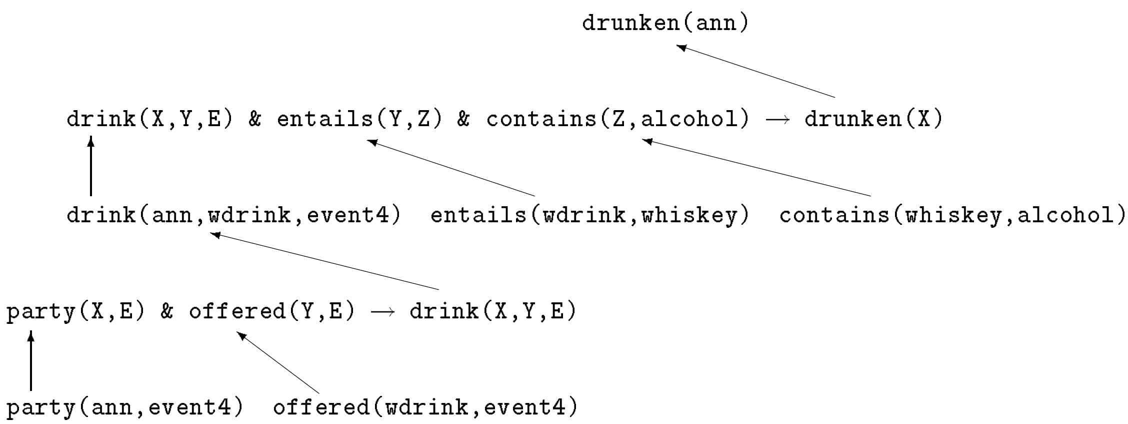


Figure 1: Derivation tree

3.2 Short overview of MOBAL

We may now give a short overview of the system which we used for the simulation. MOBAL uses a function-free first-order logic as its representation formalism. First-order logic is further restricted in that the conclusion consists of one predicate only. The well-defined semantics of the representation formalism uses four truth values: true, false, unknown and contradicting (see [Wrobel, 1994]). In addition to facts and rules as described above, integrity constraints can be represented in MOBAL. For instance,

```
covers(O1,O2,Me,E) & covers(O2,O1,Me,E) ⇝
```

expresses that from the point of view of `Me` it cannot be the case that the object `O1` covers `O2` and at the same time (`E`) the object `O2` covers `O1`.

MOBAL offers a set of tools that support the user in manually developing a knowledge base. The knowledge base is then used in order to derive new facts. The inference engine can enrich a set of facts using rules in a forward chaining manner: instantiaton of the premise of a rule by given ground facts leads then to the derivation of the instantiated conclusion. The inference engine can answer questions using rules in a backward chaining manner: the instantiation of the conclusion is propagated to the premise and each predicate of the premise is then tried to verify by applying a rule or finding a unifying ground fact.

MOBAL offers induction operators for the learning tasks of concept learning and rule discovery. Concept learning can be embedded into knowledge revision and can be used for constructive induction. A newly constructed and defined predicate can be used as additional condition in the premise of a revised rule. We have introduced the induction of concept definitions, theory revision including a way of conceptual change, and ways of changing the representation language. Note, that it is possible to model the learning of relational concepts using these learning operators in concert. For instance, we may start with predicates for absolute weight and then introduce relational predicates for weight into the representation language. Werner Emde has made experiments concerning explanations of why some objects float and others sink [Emde, 1987].

What about the representation formalism? Can it be changed, too? There exist knowledge compilation techniques that transform the complex first-order logic formulas into the simpler ones of propositional logic (see, e.g. [Lavrač and Džeroski, 1994]; [Rieger, 1997]). Processing the compiled knowledge is much faster than the originally acquired version of the knowledge. Compilation corresponds to the transition from novice to expert. The other way around, the acquisition of a more complex representation formalism on the basis of a less complex one – as far as we know – has never been formalized. If such a transition would be necessary for the simulation of conceptual change in children, current knowledge-based systems cannot be used for this task. It is an open question whether children start with strongly restricted representation capabilities (in the sens of the representation *formalism*) and acquire more expressive and difficult ones, or are born with the most complex representation capabilities and learn to apply them more and better (in the sense, that their representation *language* is enhanced). The latter position is clearly advocated by Noam Chomsky [Chomsky, 1981]. We cannot discuss this far-reaching matter here. It does not touch our study, because the acquisition of relational representations in the sens of the representation formalism takes place early in life. We safely assume that at the age of the youngest children in the empirical study (6.9 years old) children have all representation capabilities (representation formalism). Of course, they still enhance their conceptual models which may require a representation change in the sense of enhancing the representation language.

4 The Computational Model of the Day/Night Cycle

A sufficient generality has been formulated as an important demand on the formal representation of the children's explanations. This means that the formal language allows for representing everyday events of appearance and disappearance rather than be narrowed to events concerning solely the day/night cycle. This decision has an important consequence for the representation of the objects involved, be it earth, sun, and moon in the day/night cycle or wall, door, and me (my person) in everyday events. One has to decide between their representation as constants or predicates (cf. Section 3.1). In the case at hand, the objects have to be represented as constants, since their features of being stationary, opaque etc. are of primary interest and build the basis for relations to further concepts such as movements, coverage, disappearance and eventually day/night cycle. Hence, the the features of the objects have to be represented as predicates, and relations between these have to be represented by rules.

As certain characteristics of the objects involved in the day/night cycle are not directly observable, the children make assumptions about these in order to apply their concepts of appearance and disappearance to the day/night cycle. The process of making assumption (abduction) has been anticipated in the formal models by predefining appropriate sets of assumptions for the initial, synthetic and scientific models. Of course, these predefined assumptions simplify the computational models. On the other hand, in the formal representation the underlying assumptions have to be made explicit and thus disclose and question important aspects of the children's explanations. Furthermore, assumptions may be less reluctant to change by the treatment of contradictions than theoretic statements. The same distinction may hold for statements and observations. Though the MOBAL system can deal with evidence for facts and assumptions, this feature has not been applied in the first modelling for efficiency reasons.

Even if everyday experiences are in the scope of the formal language and taking into consideration that these are intertwined with explanations for the day/night cycle, there has to be a pragmatic tradeoff between an extensive representation and a well-founded representation, because the underlying psychological study naturally falls short in accounting for all the background knowledge involved in explaining the day/night cycle and not every representation is tractable in computational terms. Once again, the computational simulation is to make these drawbacks explicit for designing further investigations. However, some representational decisions had to be made, and this means that some background knowledge used by the children may not be part of the model. A similar design decision had to be made regarding the representations of spatial relations between objects. In addition, the formal representation is just one of many possible interpretations of the natural language description the children gave their explanations in.

This has to be subject to empirical verifications.

This section is organized as follows: First, the representation of the different explanations is illustrated; then its validation is shown by asking the operational models the same questions that the children have been asked, and finally these models have been systematically investigated to find possible transitions between the models that are caused by new, scientific information.

4.1 The Representation of Different Explanations

The sixteen explanations for the day/night cycle presented in [Vosniadou and Brewer, 1994] differ significantly in terms of comprehensiveness. In particular, the explanation types 4 and 12 to 16 are unspecific to such a degree, that they cannot sensibly be reconstructed (see table 1). Therefore, not all of the sixteen explanations are suited for a deep modelling of the underlying concepts. The explanation type 2 is based on rather different answers by the children [Vosniadou and Brewer, 1994], i.e., (*The sun goes down into the ground.* and *The sun goes down on the ground behind hills.*). Hence these explanations will be treated as distinct. Since the moon plays a role in only a few explanations and in combination with the sun. Other celestial bodies such as the stars do not occur in any of the explanations. In the formal models, only the sun, but not the moon and the stars have been modeled. In fact, nine explanations are sufficiently detailed and have been represented as computational models (cf. [Mühlenbrock, 1994a],[Mühlenbrock, 1994b]). These are the following (the numbers in brackets refer to the explanation types in table 1):

- Model 1: *Clouds cover the sun* (1)
- Model 2: *The sun goes down into the ground* (2)
- Model 3: *The sun moves behind hills* (2)
- Model 4: *The sun goes far away* (5)
- Model 5: *The sun moves to the other side of the earth* (3)

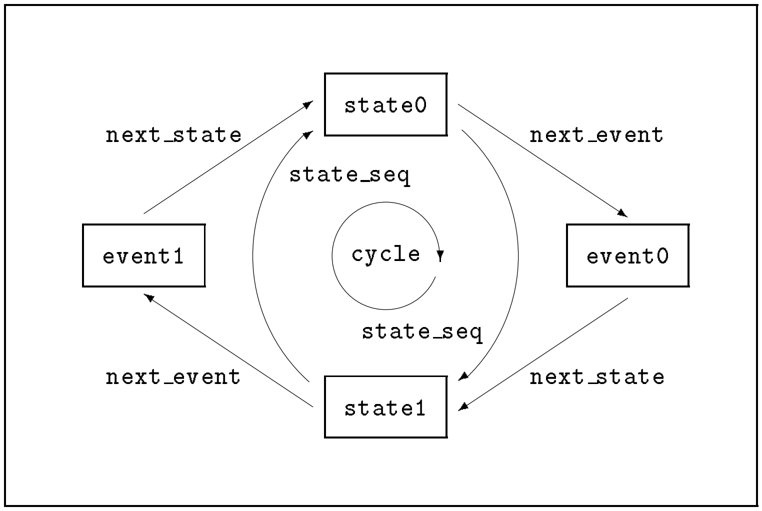


Figure 2: The temporal structure of the domain

- Model 6: *The sun revolves around the earth* (6)
- Model 7: *The earth revolves around the sun* (7)
- Model 8: *The earth turns up and down* (8, 9)
- Model 9: *The earth rotates* (10, 11)

Models 1 to 4 have been classified initial, models 5 to 8 synthetic, and model 9 represents the scientific explanation for the day/night cycle (cf. section 2).

For the modelling, the obvious objects in this domain are the sun, the earth (ground) and clouds. Less obvious, but nonetheless included in the explanations, is a designated person (me), who observes the events taking place. Theses objects have been represented as constants as described above, namely `sun`, `earth`, `cloud` and `me`. Predicates are used to represent the relevant objects' features, such as `observing`, `shining`, `opaque`, and `solid` or relations between objects such as `bigger`. In addition, the explanations implicitly contain separate points (intervals) in time where things change and where things are constant. In the day/night cycle the relevant temporal distinctions are between two states (`state0` and `state1`) and two events (`event0` and `event1`), that will be characterized as day and night or sunset and sunrise, respectively, in a valid derivation (explanation) of the day/night cycle. A temporal structure is imposed on these (temporal) objects by means of predicates such as `next_state`, `state_seq`, and `cycle` (cf. Figure 2). The spatial structure is expressed in terms like behind, other side, around, etc. Its representation is based on the general predicates `in` and `complementary` that describe relations between objects and the relevant areas (directions) regarding the objects. A sample spatial relation is shown in figure 3, which can be represented as `in(object1,area1,object2,state)` and particularly `in(sun,up,earth,state0)` in the day/night cycle domain.

These predicates together with the constants form the basic layer of the conceptual structure. More complex relations are formed by means of rules and integrity constraints (cf. section 3.2). Integrity constraints are used to prevent cycles in the rulebase. A complete list of all rules and constraints concerning the day/night cycle is given in appendix A and B,

Predicates	Arguments				
`appears`	who	for who	when		
`between`	who	between who	and who	when	
`bigger`	who	als who			
`complementary`	which	and which			
`covers`	who	who	for who	when	
`cycle`	when	before when	before when	before when	
`day`	who	when			
`day_night_cycle`	who	when	before when	before when	before when
`disappears`	who	for who	when		
`event_seq`	when	before when	when		
`hides`	who	behind who	for who	when	
`in`	who	in which	of who	when	
`infinite`	who	which			
`invisible`	who	for who	by who	when	
`moves`	who	for who	when		
`next_area`	which	before which			
`next_event`	when	before when			
`next_state`	when	before when			
`night`	for who	when			
`observing`	who				
`on`	who	on who			
`opaque`	who				
`reappears`	who	for who	between who	and when	
`returns`	who	for who	when		
`revolves_acw`	who	for who	when		
`revolves_cw`	who	for who	when		
`rotates_acw`	who	for who	when		
`rotates_cw`	who	for who	when		
`shining`	who				
`solid`	who				
`state_seq`	when	before when	before when		
`stationary`	who	when			
`uncovers`	who	who	for who	when	
`unhides`	who	behind who	for who	when	

Table 2: Specification of the predicates (in alphabetical order)

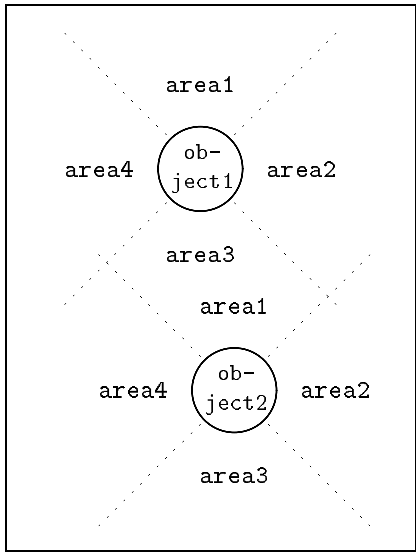


Figure 3: The local structure of the domain

respectively. On the next layer there are concepts for certain kinds of movements (`moves`, `revolves`, `rotates`) and of (dis)appearance (`covers`, `uncovers`, `hides`, `unhides`). Finally, these relate to `day` and `night` and the `day_night_cycle`. Using these concepts, the nine different models for the day/night cycle have been represented by formulating appropriate facts for each of them so that the system derives an explanation according to the specific models (see appendix C for the specific facts of the nine models).

4.2 The Validation of the Representation

We validated the computational model by asking the same questions as were asked the children and checked the system's answers. The central question "Where is the sun at night?", that is queried by

Model	`in(sun,X,Y,Z) & night(me,Z)`	"Where is the sun at night?"

1	`in(sun,up,cloud,state1).`
	`in(sun,up,earth,state1).`
	`in(sun,up,me,state1).`
		"Behind clouds."

3	`in(sun,right,hill,state1).`
	`in(sun,right,me,state1).`
	`in(sun,up,earth,state1).`
		"Behind hills."

4	`in(sun,up,earth,state1).`
	`in(sun,up,me,state1).`
	`in(sun,up,visual_range,state1).`
		"Far out into space."

2, 5, 6, 7, 8, 9	`in(sun,down,earth,state1).` `in(sun,down,me,state1).`	"Down. At the other side of the earth."

Table 3: Querying the models

```
in(sun,X,Y,Z) & night(me,Z)
```

results in different answers according to the explanation of the day/night cycle (cf. table 3): Model 1 (*Clouds cover the sun*) answers, that the sun is above the clouds, the earth and me (observer). In model 3 (*The sun goes behind hills*) the sun is also above the earth, but it is in the same direction (`right`) to the hills as the hills are to me, hence the hills are between me and the sun. Model 4 (*The sun goes far away*) says that the sun is above the earth and me and beyond the visual range. In all the other models including the scientific explanantion the sun is below the earth and below the observer, i.e. on the other side of the earth.

In the formal models, an explanation for the day/night cycle is represented as a derivation tree for the predicate `day_night_cycle`. A sample derivation is shown in Figure 4, that represents the scientific explanation for the day/night cycle. As is easily seen, an explanation is quite a complex structure. It relates general knowledge about a temporal scheme (the derivation of `cycle` in the left most part of the graphic), about the meaning of spatial relations (the derivation of `covers, uncovers` in the right most part of the graphic) and movements (in the middle of the graphic). The specific knowledge about the sun, the earth, and the observer makes rule application possible. For instance, the movement `moves(earth, sun, event1` is only possible, because the earth is assumed to be not stationary. The more specific characterisation of the movement ascribes areas to the involved objects (particularly a `down` area to the earth). It takes a lot of knowledge and inference to conclude `between(earth, sun, me, state1)`! Further inference is necessary to link this with the reappearance of the sun in the morning. The short answers of the system are based on a complex explanation.

In the sample derivation figure 4, the fact

```
day_night_cycle(me,state0,event0,state1,event1)
```

is inferred from the rotation of the earth

```
rotates_acw(earth,sun,event0)
```

and of the observer

```
rotates_acw(me,sun,event0)
```

and from the coverage of the sun by the earth

```
covers(earth,sun,me,event0)
```

which leads to the disappearance of the sun

```
disappears(sun,me,event0)
```

It is worth noting that though the observer (me) occurs in most of these facts, the explanation of the day/night cycle is not bound to just this constant `me`. In fact, another observer, e.g. `someone_else`, who stays on the other side of the earth may be added to the model by additional facts stating the relevant temporal and local relations. Then model 9 will derive that it is day for the second observer when it is night for the first one and vice versa, i.e.

```
night(me,state1), day(me,state0)

day(someone_else,state1), night(someone_else,state0)
```

This only holds for the synthetic models and the scientific model, since in the initial models it will lead to contradictions with the prevailing definition of the earth. The representation language is also general enough to describe everyday events of appearance and disappearance. Provided the relevant facts are given it will derive an explanation based on coverage in a situation where someone disappears caused by closing a door.

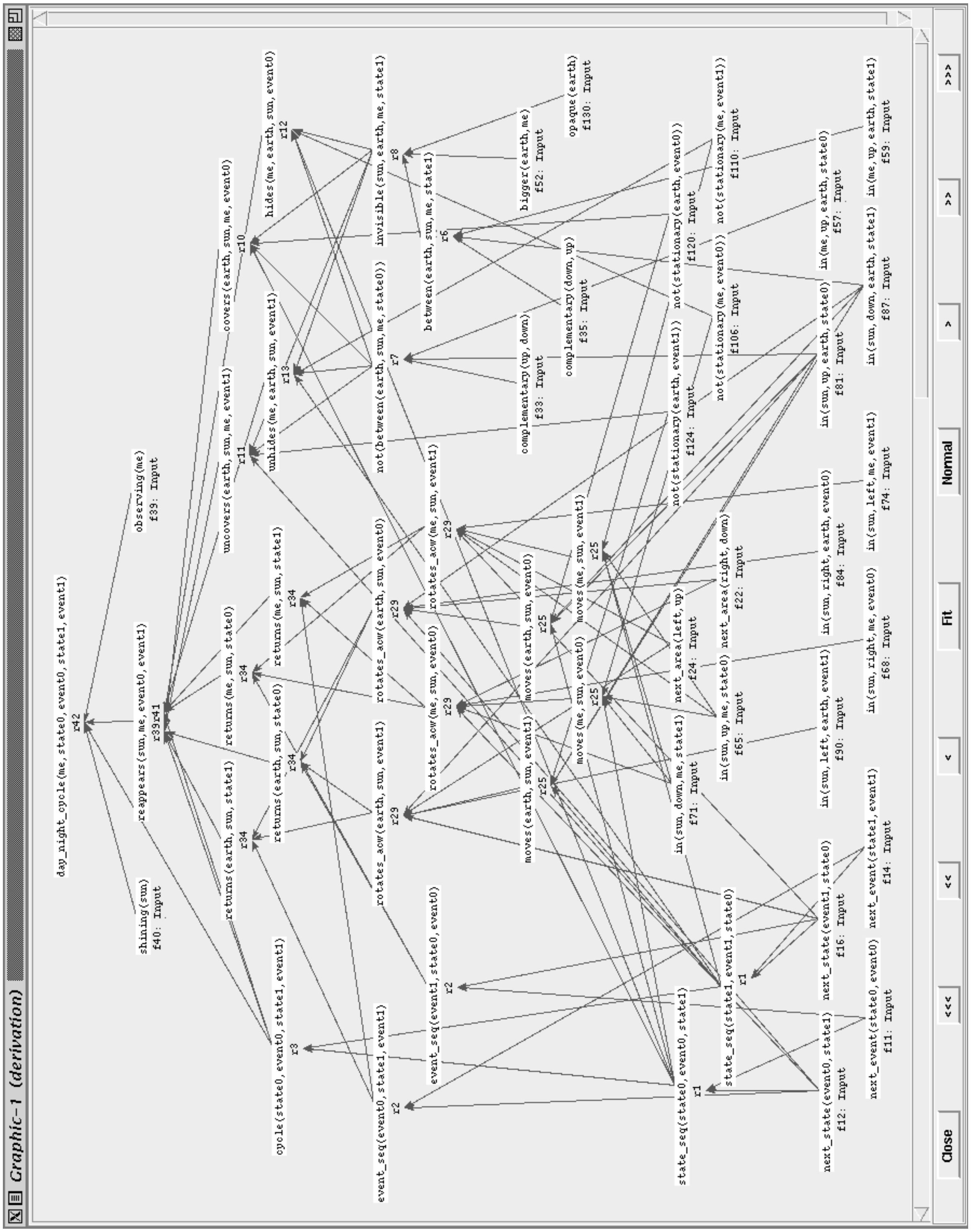


Figure 4: Derivation for `day_night_cycle` in the scientific model

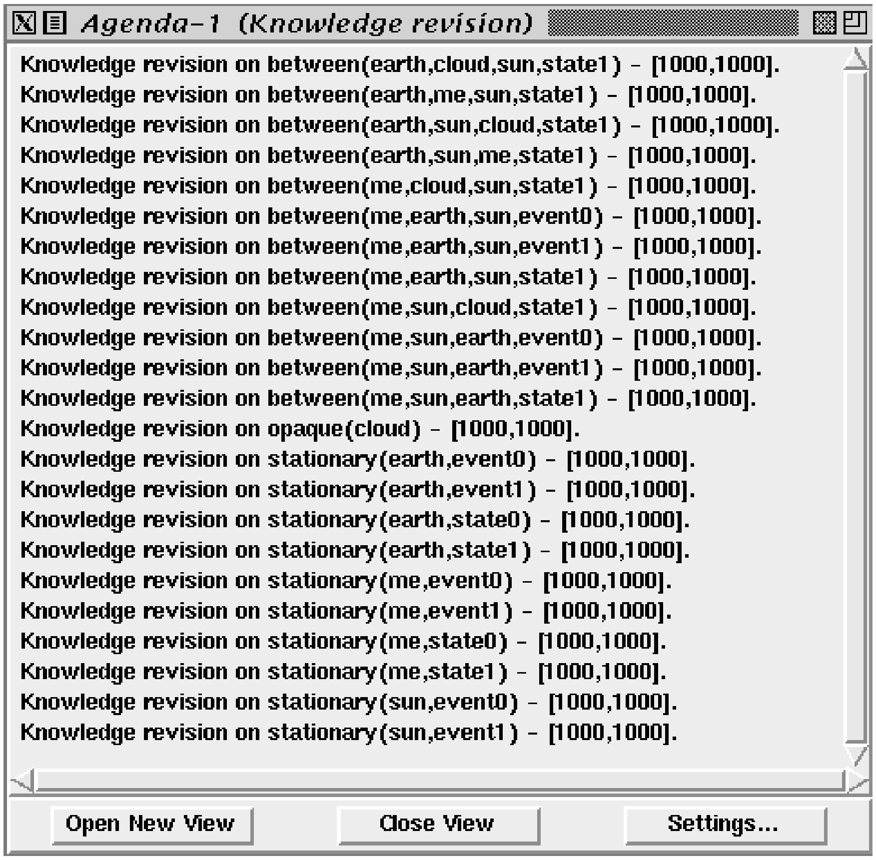


Figure 5: Contradictions between Model 1 und Model 9

4.3 Experiments

We then made experiments with the transition of one explanation to another one. Using the knowledge revision component we found for all non-scientific explanations the contradictions to the scientific explanation. Those statements of the scientific explanation, which are in conflict with the respective other explanation, are candidate instructions. We experimented with all subsets of the candidate instructions. Inputing instructions to an explanation leads to a new state of knowledge because of the revisions. Now, two cases can be distinguished. First, the revised knowledge is no longer capable of predicting that there will be a sun rise giving daylight followed by sun set and moon shine. In this case, the child will stick to the unrevised knowledge, simply ignoring the teacher's input. Second, the revised knowledge is a new explanation for the day/night cycle. We tested every explanation together with its corresponding instructions. We found out that only some of all possible transitions are possible by true instructions. For instance, to move from one naive explanation to another naive explanation, a wrong statement must be input. Hence, in Figure 6, there is no transition between the naive explanations.

The addition of scientific information of the day/night cycle to the initial and synthetic models leads to contradictions. The system places contradiction in an agenda for revision purposes. Figure 5 lists contradictions between the scientific model and model 1 (*Clouds cover the sun*). There are also violations of integrity constraints. Some scientific facts are critical for

Transitions[a]	**Facts**	**Operations**[b]
$1 \rightarrow 6$	`not(opaque(cloud)).`	–
	`in(sun,down,earth,state1).`	4 ∅ 1 $\circlearrowleft$
	`in(sun,left,earth,event0).`	2 ∅ 1 $\circlearrowleft$
	`in(sun,right,earth,event1).`	2 ∅ 1 $\circlearrowleft$
$6 \rightarrow 9$	`stationary(sun,event0).`	–
	`not(stationary(earth,event0)).`	1 ∅
	`not(stationary(earth,event1)).`	1 ∅
$1 \nrightarrow 5$	`in(sun,left,earth,event0).`	!!
	`in(sun,left,earth,event1).`	!!

[a]transitions: possible ($\rightarrow$) and impossible ($\nrightarrow$)

[b]operations: deletion (∅) and revision ($\circlearrowleft$); unscientific facts (!!) and destructive facts (–)

Table 4: Possible transitions between models 1, 6 and 9 and impossible transition between models 1 and 5

the destruction of a wrong explanation. In model 1 e.g., the addition of `not(opaque(cloud))` brings about the complete retraction of the derivation of `day_night_cycle`. When adding further scientific facts, the system asks for revisions and deletions of facts. This process eventually leads to the creation of the scientific explanantion for the day/night cycle. Intermediately, the system may derive an explanation according to model 6, depending on the sequence of input facts (cf. Figure 4).

The transition from model 1 to model 6 takes place by giving the information that the sun is on the other side of the earth at some state and that on its way passes different sides of the earth in the events beforehand and afterwards:

```
in(sun,down,earth,state1).

in(sun,left,earth,event0).

in(sun,right,earth,event1).
```

Further information about the mobility of the earth and the immobility of the sun (regarding the day/night cycle!), i.e.

```
stationary(sun,event0).

not(stationary(earth,event0)).

not(stationary(earth,event1)).
```

lets the system derive model 9. If the order of input facts is altered, the derivation of model 9 will nonetheless be achieved after all, yet model 6 will not be an intermediate explanation. On the other hand, e.g. the transition from model 1 to model 5 is not possible by adding scientific facts. For this transition, the incorrect information that the sun passes the same side of the earth before and after its movement to the other side of the earth has to be added, i.e.

```
in(sun,left,earth,event0).
```

Transitions	**1 → 6**	**6 → 9**	**1 → 9**
starting facts	295	244	295
input facts	4	3	7
treatments[a]	11	2	13
final facts	244	250	250
equal facts	215	233	210
deleted facts[b]	80	11	85
new facts[c]	29	17	40
effect[d]	2,75	0,67	1,86
deletion[e]	20	3,67	12,14
renewal[f]	7,25	5,67	5,71

[a]deletions (∅) and revisions (⭮)
[b]difference starting facts − equal facts
[c]difference final facts − equal facts
[d]Differencd treatments / input facts
[e]Quotient deleted facts / input facts
[f]Quotient new facts / input facts

Table 5: Costs of transitions

```
in(sun,left,earth,event1).
```

Hence the transition from model 1 to 5 cannot be achieved on the basis of purely scientific information.

The formal representation allows us to measure the complexity of the transitions. For each of the possible transitions we count the total number of revisions necessary and calculate the revisions per input. We found out, that the number of revisions per input are more or less the same. A more interesting point is to compare the complexity of the direct transition from a naive explanation to the scientific one with the complexity of the transition from a naive explanation to an intermediate one or from an intermediate one to the scientific one (see Table 5 for the complexity of the direct and indirect transition from explanation 1 to the scientific explanation).

Figure 6 is showing the possible transitions together with their difficulty in terms of necessary revisions. As can be seen, there are several initial/synthetic/scientific transitions , e.g. 1 → 6→ 9 and 2 → 5→ 9. There are also two direct transitions from an initial to the scientific model, namely 3 → 9 and 4 → 9. All synthetic models can be transformed into model 9, but there is no transition from an initial model to model 7. For model 7, one has to add the information that the earth is revolving around the sun. This is not correct for the day/night cycle, but it is a scientific information concerning the explanation of the alternation of seasons. Thus, the emergence of model 7 seems to be originated by mismatching information on explaining different astronomic phenomena. In other words, the revolution of the earth around the sun does take place, though it is not relevant in regards of the temporal structure of the day/night cycle (`cycle(state0,event0,state1,event1)`).

There are subsets of candidate instructions that lead to an intermediate explanation type. The overall number of revisions that have to be performed in order to reach an intermediate explanation were between 4 and 11. The number of revisions for moving from an intermediate to the scientific explanation were between 1 and 5. The number of revisions when moving

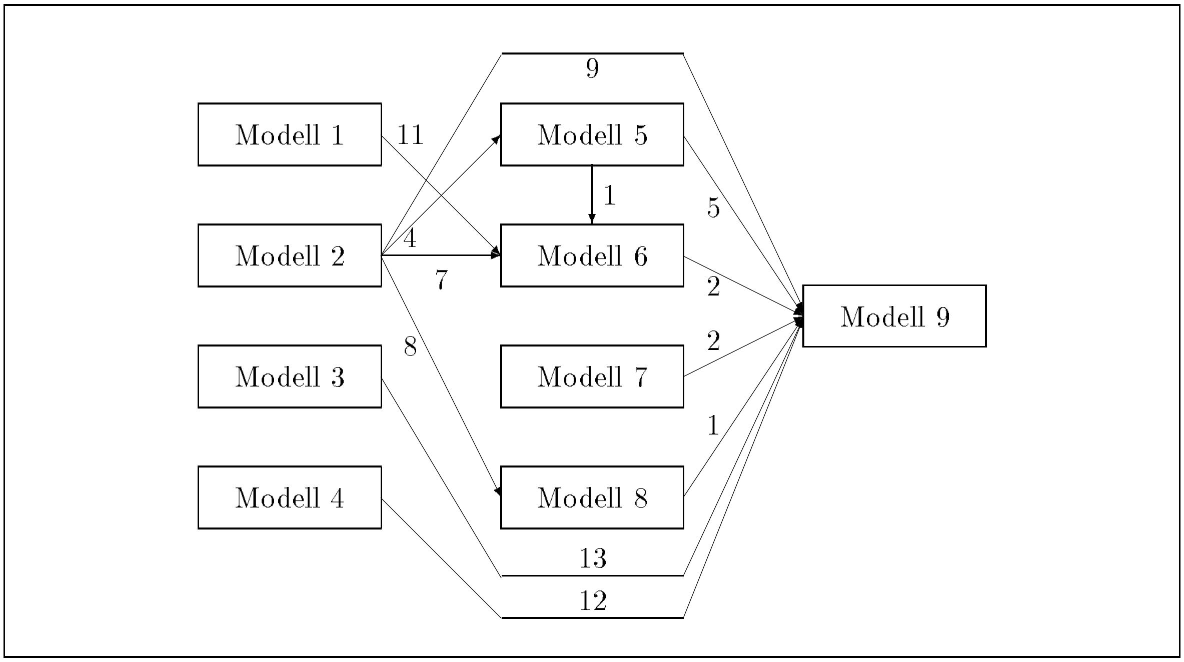


Figure 6: Possible transitions between the different models and their costs

from an initial to the scientific explanation were between 9 and 13. This does not sound very distinguished, because there is an overlap of the intervals. If we take a closer look, there is the transition from explanation 1 to explanation 6 which exceeds the threshhold of 9 revisions. For the direct transition from explanation 2 to the scientific one, only 9 revisions are necessary. Now, the question is whether there is any individual child that moved from explanation 1 (*Clouds hide the sun at night*) to explanation 6 (*The sun revolves around the earth*). The other question is, whether children believing, that the sun goes into the ground at night, learn the scientific explanation more easily than children with other initial explanations. What our computational model states is that

- not all transitions between explanations are possible;
- transitions have different degrees of difficulty.

4.4 Conclusion

We have developed a computational model of children's explanations of the day/night cycle. Since we had to represent the meaning of natural language and pictoral descriptions as reported in [Vosniadou and Brewer, 1992, Vosniadou and Brewer, 1994], there are many open questions concerning the particular representation language and formalism. More detailed empirical data are needed in order to clarify them. For instance, we wondered how the children define day and night. Does the sun make the day and the night is just not seeing the sun? Or does the sun make the day and the moon makes the night? If the moon makes the night, what happens if the moon is occluded by clouds – will this cause day? We took into account more than 20 studies on children's explanations of astronomic phenomena (see, e.g., [Nussbaum, 1989b]) and were particularly aware of the development from absolute space to relative space. However, these studies are not linked with the empirical data at hand. Hence, we could use the literature only as a guideline. It is a problem to base a detailed computational model on data that were acquired with a different goal. For instance, we wondered, how the children understood the pictograms used in the empirical inquiry, where an abstracted human being was scribbled on top of a circle representing the earth. It is unclear whether the children in fact considered this human being (represented in the computational model as `me`) themselves or other real persons. As a guideline we checked whether the same representation language is capable of expressing everyday experience of disappearing/reappearing objects. This gives necessary but not yet sufficient evidence. The resulting computational model was then tested with respect to its behavior as compared to the answers of children as reported in [Vosniadou and Brewer, 1994]. The computational model describes each of the explanation types of the children such that the system gives the same answers as did the children (assuming a clear relation between the natural language answers and the ones in formal language). This means that the computational model in fact expresses the empirical data. Using the picture from the introduction, we may state that the landmarks have been met.

Transitions from one explanation to another one have been modeled. In contrast to most other computational investigations of conceptual change, we have investigated *sequences of transitions*. The experiments explore paths between landmarks and describe in which order under which assumptions landmarks can be reached with which difficulty. The transitions were triggered by true input facts which contradict the current explanation ot the day/night cycle. Using true input facts, not all transitions are possible. In other words, some transitions are only possible using false facts. In particular, the transition from one naive explanation to another naive explanation requires false assumptions. It is also interesting that using true facts no transition from any of the naive explanations leads to the explanation that the earth revolves around the sun. We suspect the child who gave this explanation to parrot a true statement without having understood it, i.e. without embedding it into its conceptual structure.

The process of transition from one theory to another could effectively be modeled using the knowledge revision operator. The computational model does not take the perspective of the expert and describes all *we* know about the day/night cycle. In contrast, within the computational model, all explanations are consistent – until an input fact introduces a conflict. Each formally represented explanation is a theory in its own right and only this is used to determine a contradiction which is triggered by an input fact. Interestingly enough, revision did not delete all facts from a previous explanation. Hence, some facts from former explanations remained in the current explanation. Practically, this made necessary a program that determines the explanation type of the current explanation. Interpreting it psychologically, this means that statements from former explanations remain in the current knowledge state. The computational model precisely seperates between facts that are to be deleted in the course of knowledge revision, and those that remain untouched. Can this be empirically validated?

In contrast to our expectations, no constructive induction step – embedded into knowledge revision – was necessary. No representation change according to our definition was necessary or even useful. All rules remained unchanged. Recall that we have represented spatial and temporal reasoning in the form of rules. The rules represent the meaning of basic concepts of disappearing and related spatio-temporal processes as they are used in everyday-life. These basic concepts are unchanged in the course of the evolving understanding of the day/night cycle. Basic concepts can be acquired using inductive learning operators. They can be instantiated by a coin disappearing in the sand or by a stone disappearing in the sea as well as by the sun disappearing in the infinite earth. The facts represent features of objects and their relations. These determine whether an object consistently instantiates a particular disappearing process or whether the instantiation leads to a contradiction. The system MOBAL expresses inference depth, which means the number of inference steps the system makes. If a contradiction lies outside the range of this inference depth it is not detected. If there were data available, the system could predict behavior under certain contextual settings, e.g., a current problem solving activity, questions that have to be answered, available time for a task, and other influential factors. In our experiments, changes of the knowledge state all involved assumptions about features and relations of the earth, the sun, hills and clouds. Triggered by new facts, the assumptions – mainly about the earth – were revised. This led to new derivations and hence to a new explanation of the day/night cycle. Can this be generalized? The question is whether the set of inferential procedures and the basic representation language is acquired until the age of about 6 years. Learning from then on becomes the acquisition of more knowledge about objects, their properties and relations between them, together with the revision of the current knowledge state. This question requires further interdisciplinary work.

There is another conclusion we could draw. The earth and the sun are not concepts but individual objects. The corresponding concepts (e.g., planets, galaxies) have not been investigated in the empirical investigation and the corresponding simulation. In the computational model, the earth and the sun are not represented as instances of a known concept at all. Even in the representation of the scientific explanation, there is no need to classify earth and sun as celestial objects or planets. It makes a great difference, whether something is an individual object (a constant argument) or a defined class of objects (a predicate defined by a rule). Therefore, it could well be that the day/night cycle explanations do not require representational change, because assertions about individual objects play the most important role. An empirical investigation and the computational modeling of its results could shed more light on this matter: are other changes than revision, in particular constructive inductive learning operators, necessary in order to model the acquisition of knowledge about galaxies and stars?

Characterizing the possible paths to the landmark of current scientific knowledge, we used complexity measures of the revision process. We found that the one-step transition from a naive to the scientific explanation is more complex than the step from a naive to an intermediate explanation and the step from an intermediate explanation to the scientific one. We have measured the difficulty of the transition from one explanation to another one in terms of the number of necessary revisions. Does this correspond to a threshhold of complexity that can be observed in human thinking? This question could be answered by the empirical investigation of other domains and the computational modeling of these results.

5 Discussion from a Psychological Point of View by Stella Vosniadou, Univ. Athens

Let me start this discussion by saying that Morik and Mühlenbrock have moved a long way towards dimystifying representational change for machine learning. Their work is of great interest for its originality as well as for the fact that the system they developed succeeds in simulating children's explanations of the day/night cycle. As we move from dimystification to science (see Van Lehn, this volume), it becomes critical to examine what has been achieved in this work both from the computational and from the psychological point of view and to set the ground for future tasks and collaborations. The present discussion presents the point of view of the psychologist.

One important result of the research presented in this paper is the finding that no representation change (as it is defined in the paper) is necessary to model the transition from the initial explanations of the day/night cycle to the "scientific" explanations. It may turn out that Morik and Mühlenbrok are right in coming to this conclusion. Nevertheless, at this point in their research, this claim should not be left as such without further examination.

The validity of the claim that no representational change is needed to account for the transition from the initial, nave models of the day/night cycle, to the models that we consider "scientific" is necessarily tied to the specific representation language used to express children's initial models. However, there are some decisions that have been made about the conceptual system of the young child that require further investigation. One such decision is the adoption of a complex temporal and spatial conceptual structure to embed the young child's representation of the earth, something that makes it possible for the learning system to go from a flat earth model to a spherical earth through revision mechanisms only. The empirical research has shown that the transition from the flat earth to the spherical earth model is the most difficult transition on the way to creating a "scientific" model of the day/night cycle.

Another questionable assumption is that young children are able to distinguish between their point of view and that of the earth's. Finally, it is not entirely clear why predicates such as rotates and revolves, need to be assumed in the conceptual system of the young child, when a more general predicate like "moves" could be adequate to represent the initial mental models we have identified.

Morik and Mühlenbrook claim that the transition from one initial model to another initial model requires false facts. This is an interesting finding, consistent with the Vosniadou and Brewer (1994) assumption that there is no transition from one initial model to another. However, recent findings have shown that and it is possible that some initial models may co-exist without being contradictory in the conceptual system of the young child. (See the Vosniadou et al paper in this volume).

There is a more general question here and it has to do with how a system decides what is contradictory and what is not. In our work we find that between 15means models that are internally inconsistent. We believe that this may happen for two reasons. One is that information that appears contradictory from an expert point of view may not be so for a child or an adult novice, because of the differences in the underlying conceptual organization. In some cases what may appear contradictory from the point of view of the expert is simply not so from the point of view of the novice.

In some other cases, children or novices, may understand that they are contradicting themselves only when the contradiction is pointed out to them, but they may not be able to discover the contradiction on their own. This is particularly true for young children, who often lack the metaconceptual awareness needed to become aware of certain contraditions. Metaconceptual awareness (i.e., awareness of one's own beliefs and presuppositions) is an important development in the conceptual system of the young child that may require representational change.

Finally, even when students are aware of contradictions they often patch them up in ways that may lead to synthetic models and misconceptions. An interesting research question for Machine Learning would be to investigate to which extend learning systems such as the above could represent learning with and without this capability.

In real life it is possible to think of situations where "true facts" can create "false assumptions". In cases such as the day/night cycle, it is possible to have linguistic inputs that are "culturally correct" without being "scientifically correct". Statements such as "the sun went behind the mountains", "the sun set", "the moon rose" etc., are common in everyday language, particularly the language addressed to young children. Such statements may reinforce the kinds of initial explanations of the day/night cycle that we find in our research.

Furthermore, it is often the case that students will misinterpret scientific terms used correctly in instruction. Let us not forget that there can be a great deal of ambiguity in the use of such terms since they often have different meanings in everyday language as compared to science (e.g., think for example of the word "force"). It would be interesting to think about how "true facts" can lead machines to "false assumptions" as they do in humans.

Before ending we should mention that it is an intrinsic limitation of machine learning systems such as the one used in the present chapter that they receive information only in the form of explicit, linguistic expressions of facts. A great deal of human learning, including the initial mental models that young children build, is based mostly on observational evidence. It is important, from a cognitive science point of view, to think further about the status of empirical knowledge received through perception as compared to information received linguistically.

A Rules

```
next_event(S1,E) & next_state(E,S2) → state_seq(S1,E,S2).

next_state(E1,S) & next_event(S,E2) → event_seq(E1,S,E2).

state_seq(S1,E1,S2) & state_seq(S2,E2,S1) → cycle(S1,E1,S2,E2).

infinite(O,A) & next_event(S,E) → in(O,A,O,E).

infinite(O,A) & next_event(S,E) → in(O,A,O,S).

in(O1,A1,O2,S) & in(O3,A2,O2,S) & complementary(A1,A2) & ne(O1,O2) & ne(O3,O2) →
    between(O2,O1,O3,S).

in(O1,A1,O2,S) & in(O3,A2,O2,S) & complementary(A1,A3) & ne(O1,O3) & ne(A2,A3) →
    not(between(O2,O1,O3,S)).

between(O1,O2,O3,S) & opaque(O1) & bigger(O1,O3) → invisible(O2,O1,O3,S).

between(O1,O2,O3,S) & opaque(O1) & bigger(O1,O3) → invisible(O3,O1,O2,S).

not(between(O1,O2,O3,S1)) & invisible(O2,O1,O3,S2) & state_seq(S1,E,S2) &
    not(stationary(O1,E)) → covers(O1,O2,O3,E).

invisible(O2,O1,O3,S1) & not(between(O1,O2,O3,S2)) & state_seq(S1,E,S2) &
    not(stationary(O1,E)) → uncovers(O1,O2,O3,E).

not(between(O1,O2,O3,S1)) & invisible(O2,O1,O3,S2) & state_seq(S1,E,S2) &
    not(stationary(O3,E)) → hides(O3,O1,O2,E).

invisible(O2,O1,O3,S1) & not(between(O1,O2,O3,S2)) & state_seq(S1,E,S2) &
    not(stationary(O3,E)) → unhides(O3,O1,O2,E).

covers(O1,O2,O3,E) → disappears(O2,O3,E).

covers(O1,O2,O3,E) → disappears(O3,O2,E).

hides(O1,O2,O3,E) → disappears(O1,O3,E).

hides(O1,O2,O3,E) → disappears(O3,O1,E).

uncovers(O1,O2,O3,E) → appears(O3,O2,E).

uncovers(O1,O2,O3,E) → appears(O2,O3,E).

unhides(O1,O2,O3,E) → appears(O1,O3,E).

unhides(O1,O2,O3,E) → appears(O3,O1,E).

disappears(O1,O2,E) & next_state(E,S) & shining(O1) & observing(O2) → night(O2,S).

appears(O1,O2,E) & next_state(E,S) & shining(O1) & observing(O2) → day(O2,S).

in(O1,A1,O2,S1) & in(O1,A2,O2,S2) & state_seq(S1,E,S2) & ne(A1,A2) &
    not(stationary(O1,E)) → moves(O1,O2,E).

in(O1,A1,O2,S1) & in(O1,A2,O2,S2) & state_seq(S1,E,S2) & ne(A1,A2) &
    not(stationary(O2,E)) → moves(O2,O1,E).
```

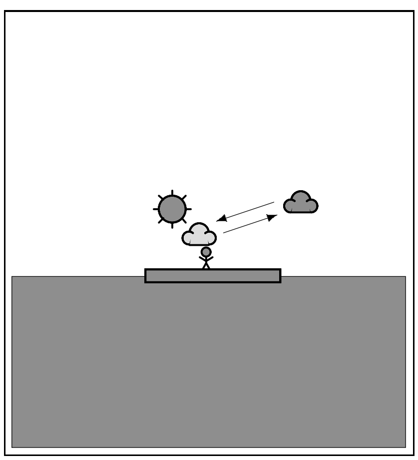
```
moves(O1,O2,E) & next_state(E,S2) & in(O1,A1,O2,E) & in(O1,A2,O2,S2) &
    next_area(A1,A2) → revolves_cw(O1,O2,E).

moves(O1,O2,E) & next_state(E,S2) & in(O1,A1,O2,E) & in(O1,A2,O2,S2) &
    next_area(A2,A1) → revolves_acw(O1,O2,E).

moves(O1,O2,E) & next_state(E,S2) & in(O2,A1,O1,E) & in(O2,A2,O1,S2) &
    next_area(A2,A1) → rotates_cw(O1,O2,E).

moves(O1,O2,E) & next_state(E,S2) & in(O2,A1,O1,E) & in(O2,A2,O1,S2) &
    next_area(A1,A2) → rotates_acw(O1,O2,E).

moves(O1,O2,E1) & moves(O1,O2,E2) & event_seq(E1,S,E2) & stationary(O1,S) →
    returns(O1,O2,S).

revolves_cw(O1,O2,E1) & revolves_cw(O1,O2,E2) & event_seq(E1,S,E2) →
    returns(O1,O2,S).

revolves_acw(O1,O2,E1) & revolves_acw(O1,O2,E2) & event_seq(E1,S,E2) →
    returns(O1,O2,S).

rotates_cw(O1,O2,E1) & rotates_cw(O1,O2,E2) & event_seq(E1,S,E2) →
    returns(O1,O2,S).

rotates_acw(O1,O2,E1) & rotates_acw(O1,O2,E2) & event_seq(E1,S,E2) →
    returns(O1,O2,S).

revolves_cw(O1,O2,E1) & revolves_acw(O1,O2,E2) & event_seq(E1,S,E2) &
    stationary(O1,S) → returns(O1,O2,S).

revolves_acw(O1,O2,E1) & revolves_cw(O1,O2,E2) & event_seq(E1,S,E2) &
    stationary(O1,S) → returns(O1,O2,S).

rotates_cw(O1,O2,E1) & rotates_acw(O1,O2,E2) & event_seq(E1,S,E2) & stationary(O1,S)
    → returns(O1,O2,S).

rotates_acw(O1,O2,E1) & rotates_cw(O1,O2,E2) & event_seq(E1,S,E2) & stationary(O1,S)
    → returns(O1,O2,S).

covers(O1,O2,O3,E1) & returns(O1,O2,S2) & uncovers(O1,O2,O3,E2) & returns(O1,O2,S1)
    & cycle(S1,E1,S2,E2) → reappears(O2,O3,E1,E2).

hides(O1,O2,O3,E1) & returns(O1,O2,S2) & unhides(O1,O2,O3,E2) & returns(O1,O2,S1) &
    cycle(S1,E1,S2,E2) → reappears(O1,O3,E1,E2).

hides(O1,O2,O3,E1) & returns(O1,O3,S2) & unhides(O1,O2,O3,E2) & returns(O1,O3,S1) &
    cycle(S1,E1,S2,E2) → reappears(O3,O1,E1,E2).

reappears(O1,O2,E1,E2) & cycle(S1,E1,S2,E2) & shining(O1) & observing(O2) →
    day_night_cycle(O2,S1,E1,S2,E2).
```

B Constraints

```
in(O1,A1,O2,T) & in(O1,A2,O2,T) & ne(O1,O2) & ne(A1,A2) ⤳ .

in(O1,A,O1,T) & in(O2,A,O1,T) & solid(O1) & ne(O2,O1) ⤳ .

stationary(O,E) & next_state(E,S) & not(stationary(O,S)) ⤳ .

not(stationary(O,S)) & next_event(S,E) & stationary(O,E) ⤳ .

stationary(O1,T) & not(stationary(O2,T)) & on(O2,O1) ⤳ .

not(stationary(O1,T)) & stationary(O2,T) & on(O2,O1) ⤳ .

in(O,A,O,T) & not(stationary(O,T)) ⤳ .

not(in(O1,A,O2,S1)) & in(O1,A,O2,S2) & state_seq(S1,E,S2) & stationary(O1,E) &
    stationary(O2,E) ⤳ .

in(O1,A1,O2,T) & in(O3,A2,O2,T) & complementary(A1,A2) & in(O2,A3,O1,T) &
    in(O3,A4,O1,T) & complementary(A3,A4) & ne(O1,O2) & ne(O2,O3) ⤳ .

observing(O) ⤳ (night(O,S)).

observing(O) ⤳ (day(O,S)).

observing(O) ⤳ (day_night_cycle(O,S1,E1,S2,E2)).
```

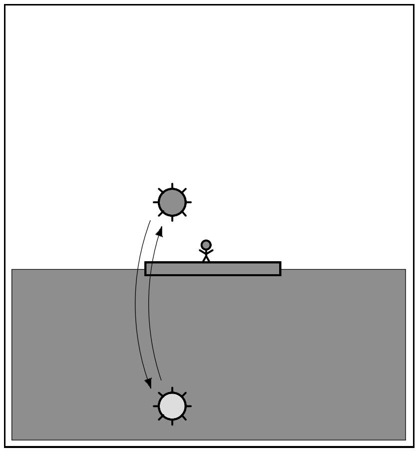
C Models and specific facts

The models and their specific facts are shown in figure 7 to 15.



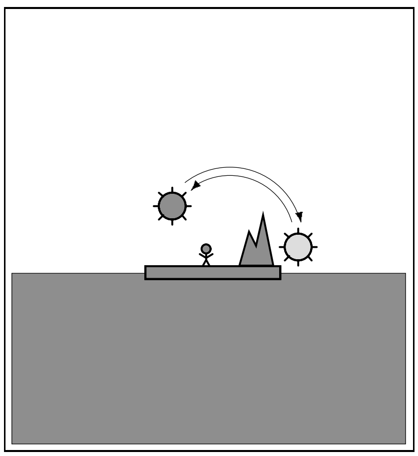
```
infinite(earth,down).
infinite(earth,left).
infinite(earth,right).

covers(cloud,sun,me,event0).
hides(sun,cloud,me,event0).
moves(cloud,sun,event0).
uncovers(cloud,sun,me,event1).
unhides(sun,cloud,me,event1).
moves(cloud,sun,event1).
```

Figure 7: Model 1: *Clouds cover the sun* (initial)



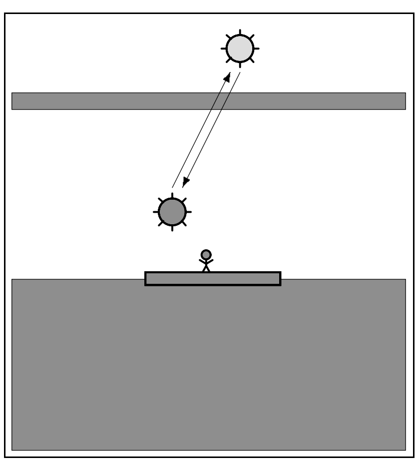
```
infinite(earth,down).
infinite(earth,left).
infinite(earth,right).

hides(sun,earth,me,event0).
revolves_acw(sun,earth,event0).
unhides(sun,earth,me,event1).
revolves_cw(sun,earth,event1).
```

Figure 8: Model 2: *The sun goes down into the ground* (initial)



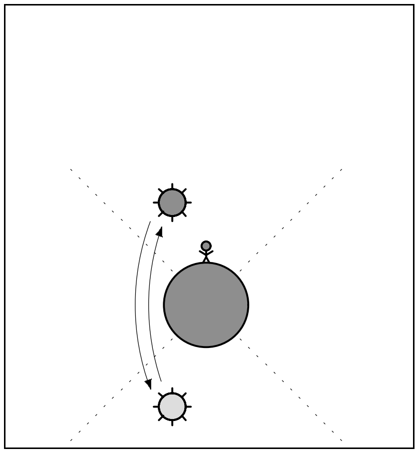
```
infinite(earth,down).
infinite(earth,left).
infinite(earth,right).

hides(sun,hill,me,event0).
revolves_cw(sun,hill,event0).
unhides(sun,hill,me,event1).
revolves_acw(sun,hill,event1).
```

Figure 9: Model 3: *The sun moves behind hills* (initial)



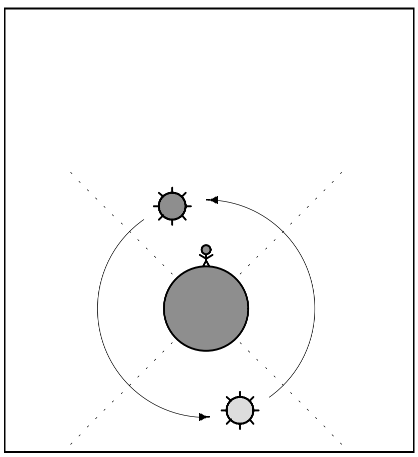
```
infinite(earth,down).
infinite(earth,left).
infinite(earth,right).

hides(sun,visual_range,me,event0).
moves(sun,visual_range,event0).
unhides(sun,visual_range,me,event1).
moves(sun,visual_range,event1).
```

Figure 10: Model 4: *The sun goes far away* (initial)



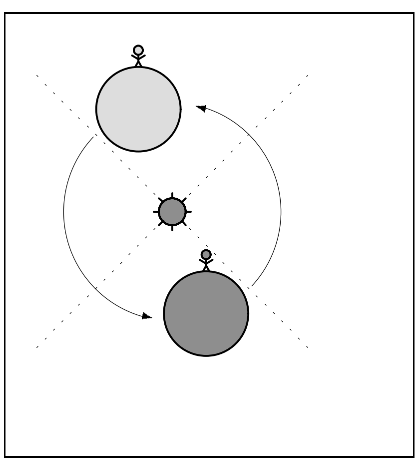
```
hides(sun,earth,me,event0).
revolves_acw(sun,earth,event0).
unhides(sun,earth,me,event1).
revolves_cw(sun,earth,event1).
```

Figure 11: Model 5: *The sun moves to the other side of the earth* (synthetic)



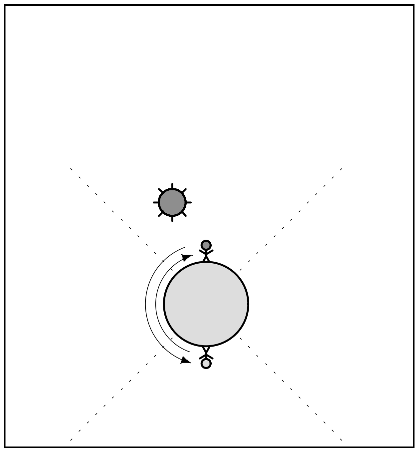
```
hides(sun,earth,me,event0).
revolves_acw(sun,earth,event0).
unhides(sun,earth,me,event1).
revolves_acw(sun,earth,event1).
```

Figure 12: Model 6: *The sun revolves around the earth* (synthetic)



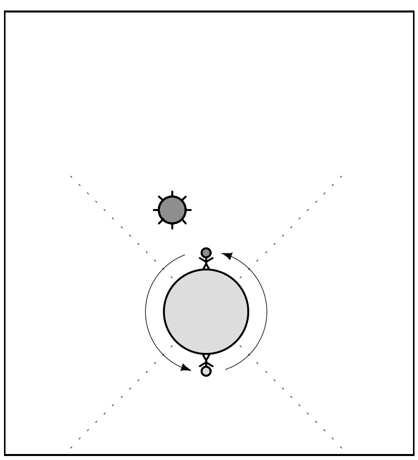
```
covers(earth,sun,me,event0).
hides(me,earth,sun,event0).
revolves_acw(earth,sun,event0).
rotates_cw(earth,sun,event0).
revolves_acw(me,sun,event0).
rotates_cw(me,sun,event0).
uncovers(earth,sun,me,event1).
unhides(me,earth,sun,event1).
revolves_acw(earth,sun,event1).
rotates_cw(earth,sun,event1).
revolves_acw(me,sun,event1).
rotates_cw(me,sun,event1).
```

Figure 13: Model 7: *The earth revolves around the sun* (synthetic)



```
covers(earth,sun,me,event0).
hides(me,earth,sun,event0).
rotates_acw(earth,sun,event0).
rotates_acw(me,sun,event0).
uncovers(earth,sun,me,event1).
unhides(me,earth,sun,event1).
rotates_cw(earth,sun,event1).
rotates_cw(me,sun,event1).
```

Figure 14: Model 8: *The earth turns up and down* (synthetic)



```
covers(earth,sun,me,event0).
hides(me,earth,sun,event0).
rotates_acw(earth,sun,event0).
rotates_acw(me,sun,event0).
uncovers(earth,sun,me,event1).
unhides(me,earth,sun,event1).
rotates_acw(earth,sun,event1).
rotates_acw(me,sun,event1).
```

Figure 15: Model 9: *The earth rotates* (scientific)

References

[Baxter, 1989] Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, 11:502–513.

[Chomsky, 1981] Chomsky, N. (1981). *Rules and Representations.* deutsche "Ubersetzung. Suhrkamp, Frankfurt a.M.

[Emde, 1987] Emde, W. (1987). Big flood in the blocks world (or non-cumulative learning). In Boulay, B., Hogg, D., and Steels, L., editors, *Advances in Artificial Intelligence II*, pages 103–109. Elsevier Pub. (North Holland), Amsterdam. also appeared in Procs. of 7th ECAI-86, Brighton, England.

[Gärdenfors, 1988] Gärdenfors, P. (1988). *Knowledge in Flux — Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, MA.

[Klein, 1982] Klein, C. A. (1982). Children's concepts of the earth and the sun: A cross cultural study. *Science Education*, 65(1):95–107.

[Klingspor et al., 1997] Klingspor, V., Demiris, J., and Kaiser, M. (1997). Human-robot-communication and machine learning. *Applied Artificial Intelligence*, 11(7/8):719–746.

[Lavrač and Džeroski, 1994] Lavrač, N. and Džeroski, S. (1994). *Inductive Logic Programming — Techniques and Applications.* Number 148 in Artificial Intelligence. Ellis Horwood, New York.

[Morik, 1996] Morik, K. (1996). A developmental case study on sequential learning: The day-night cycle. In Reimann, P. and Spada, H., editors, *Learning in Humans and Machines - Towards an Interdisciplinary Learning Science*, chapter 12, pages 212 – 227. Elsevier.

[Morik et al., 1993] Morik, K., Wrobel, S., Kietz, J.-U., and Emde, W. (1993). *Knowledge Acquisition and Machine Learning - Theory, Methods, and Applications.* Academic Press, London.

[Mühlenbrock, 1994a] Mühlenbrock, M. (1994a). Computational models of learning in astronomy. Research Report 11, University of Dortmund, Lehrstuhl Informatik VIII.

[Mühlenbrock, 1994b] Mühlenbrock, M. (1994b). Menschliches und Maschinelles Lernen – Bildung und Revision von Begriffsstrukturen zum Tag/Nachtzyklus. Diplomarbeit, Universität Dortmund, Fachbereich Informatik, Lehrstuhl VIII.

[Murphy and Medin, 1985] Murphy, G. L. and Medin, D. L. (July 1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3):289–316.

[Nussbaum, 1989a] Nussbaum, J. (1989a). Classroom conceptual change: Philosophical perspectives. *International Journal of Science Education*, 11:530–540.

[Nussbaum, 1989b] Nussbaum, J. (1989b). The earth as a cosmic body. In Driver, R., Guesne, E., and Tiberghien, A., editors, *Children's Ideas in Science*, chapter 9, pages 170 – 192. Open University Press.

[Rieger, 1996] Rieger, A. (1996). Learning to guide a robot via perceptions. In Ghallab, M. and Milani, A., editors, *New Directions in AI Planning*, pages 383–394. IOS Press, Amsterdam, Netherlands.

[Rieger, 1997] Rieger, A. (1997). *Program Optimization for Temporal Reasoning within a Logic Programming Framework.* PhD thesis, Univ. Dortmund.

[Sadler, 1987] Sadler, P. M. (1987). Misconceptions in astronomy. In Novak, J. D., editor, *Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*, volume III, pages 422–424, Ithaca, NY. Cornell University.

[Shepard et al., 1961] Shepard, R. N., Hovland, C. I., and Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75(517).

[Siegler, 1991] Siegler, R. S. (1991). *Childrens's Thinking.* Prentice-Hall, Englewood Cliffs, NJ, 2nd edition.

[Viglietta, 1986] Viglietta, M. L. (1986). Earth, sky and motion: Some questions to identify pupils ideas. In Hunt, J. J., editor, *GIREP-86:Cosmos – An Educational Challange*, pages 369–370, Nordwijk. European Space Agency.

[Vosniadou and Brewer, 1992] Vosniadou, S. and Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24:535–585.

[Vosniadou and Brewer, 1993] Vosniadou, S. and Brewer, W. (to appear 1993). Mental models of the day/night cycle. *Cognitive Science.*

[Vosniadou and Brewer, 1994] Vosniadou, S. and Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, 18:123–183.

[Wrobel, 1994] Wrobel, S. (1994). *Concept Formation and Knowledge Revision.* Kluwer Academic Publishers, Dordrecht.