



A New Mechanism for Transfer Between Conceptual Domains in Scientific Discovery and Education

Antoine Cornuéjols, Andrée Tiberghien, Gérard Collet

► To cite this version:

Antoine Cornuéjols, Andrée Tiberghien, Gérard Collet. A New Mechanism for Transfer Between Conceptual Domains in Scientific Discovery and Education. *Foundations of Science*, 2000, 5 (2), pp.129-155. 10.1023/a:1011398802965 . hal-02480323

HAL Id: hal-02480323

<https://hal.science/hal-02480323>

Submitted on 15 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A new mechanism for transfer between conceptual domains in scientific discovery and education

Antoine Cornuéjols

Equipe Inférence et Apprentissage
Laboratoire de Recherche en Informatique (LRI)
Université de Paris-Sud, Orsay
91405 Orsay Cedex
(France)
email : antoine@lri.fr

Andrée Tiberghien

UMR GRIC-COAST
CNRS - Université Lyon2
5, Avenue Pierre Mendès France
69976 Bron Cedex 11
(France)
email : Andree.Tiberghien@univ-lyon2.fr,
gcollet@ac-grenoble.fr

Gérard Collet

Abstract

Confronted with problems or situations that do not yield to known theories and world views, the scientists and students are alike. Rarely are they able to directly build a model or a theory thereof. Rather, they must find ways to make sense of the circumstances using their current knowledge and adjusting what needs be in the process.

This way of thinking, using past ways of perceiving the physical world to build new ones, does not follow a logical path and cannot be described as theory revision. Likewise, in many situations it is awkward, indeed often impossible, to resort to analogical reasoning to account for it. This paper presents a new mechanism, called 'tunnel effect', that may explain, in part, how scientists and students reason while constructing a new conceptual domain. 'Tunnel effect' is also contrasted with analogical reasoning.

1 Introduction

Ask for the first time a student to give an account of some experimental setting in terms of energy transformations. Observe the scientist trying to make sense of some baffling phenomena that do not yield to current theories. In both cases, to achieve the goal it is necessary foremost to discover a system of conceptual entities such that it becomes possible to interpret the situation by relating perceived features to these primitives, and then to make predictions and solve problems within the new theory. In both cases, it is rare, if it ever happens, that the new theory arises from scratch. Rather, the cognitive agent must rely on previous knowledge to perform the task. How is it that, sometimes with only one experiment acting as a trigger, it is possible to construct a new conceptual domain, like thermodynamics or electromagnetism, enabling to make sense of a wide range of phenomena, from a knowledge state that a priori is unsuited ? How is scientific discovery possible ? Or, in a more mundane way, how is it possible to learn new ways to interpret the world ?

These questions are central for understanding learning, helping teaching and possibly assisting scientists. Yet, they have so far received only very partial answers. Philosophers, on the whole, prefer to study how theories can and must stand the test of experiments. Cognitive scientists, specially in machine learning, have centered their works on induction, which is of little help in that matter, devoting scant attention to the fundamental problem of interpretation, analogy making, and other mechanisms that deal with making sense of the world and allowing to take advantage of multiple sources of knowledge.

This paper reports on a multidisciplinary head-on approach to the problem of learning new ways to interpret the world by relying on (and relating to) old ones. By studying how high school students address problems in conceptual domains that are new to them, we were led to analyze mechanisms that seemed to be at play in their segmenting the world, and constructing models of the situation, as well as the (re)conceptualization efforts that —sometimes— followed. In this paper, we focus on a reasoning mechanism that we hypothesize does explain the students behavior. We call it 'tunnel effect' for reasons that will be clarified later on. Like analogy, this mechanism allows the transfer of information from one conceptual domain to another one. Unlike analogy however, it does so without having to resort to two situations or cases, but only considers the one at hand, and it does not necessitate to specify beforehand a hierarchy of representation primitives in both domains (one being mostly unknown), nor to define how similarity between the two represented cases must be computed. In fact, it appears so natural that its scope covers a wide range of situations from metaphorical thinking to scientific discovery. Indeed it is possible to reanalyze scientific discoveries, like Maxwell's discovery of electromagnetism and of the ever since powerful concept of field theory, in a much simpler way than previous ones (e.g. (Nersessian, 1992)).

In the following, we start (section 2) by examining the problem at hand and how the traditional inferencing mechanisms deal with it. We then describe our experiments with students, our findings, and how they can be analyzed. The new reasoning mechanism we call tunnel effect is presented in section 3, and we show how it solves the problem defined in section 2. Section 4 contrasts tunnel effect with analogical reasoning underlying similarities and differences. Finally, section 5 is devoted to ongoing work and perspectives, as well as possible implications for educational sciences.

2 Definition of the problem : how to build new from old

2.1 Definition of the problem

To recap, the overall problem studied here can be described as the one of learning new ways of interpreting the world, making predictions and solving problems, when known conceptual universes turn out to be unsatisfactory.

Two aspects of this definition must be stressed at once. First, even though the learning agent does not know, at time t , how to interpret the environment satisfactorily, he/she still must have some means, through his/her current knowledge to perceive and describe it. Otherwise, there is simply no possibility for the agent to even become aware of a problem. Second, if we admit that the agent is able to built some interpretation of the situations at hand, as we argued is necessary, then it follows that the whole problem rests on the satisfaction criteria : how is an agent to be satisfied, or dissatisfied, with one's interpretation ? How is it that satisfaction criteria may be

defined a priori such that they specify what should be an interpretation in a yet to be discovered conceptual domain ?

Upon examination, in order to be satisfactory, an interpretation or a model¹ should :

- (1) - allow to make predictions about the world which indeed correspond to what is observed,
 - solve problems about the potential set of phenomena studied, including providing explanations for them.
- (2) - satisfy a set of "target constraints" that include deep beliefs akin to the themata of Holton (1973) and preconceptions about the world.

Point (1) above corresponds to conditions of *adequacy to the world*. Point (2), *target constraints*, to conditions that result from the previous history of the agent, which includes social and cultural biases, as well as more scientific ones. For instance, while studying black body radiation, Planck was of course concerned that his theory would fit the experimental data (criterion of adequacy to the world), but also that it would give a picture of the world, which for him was deeply related to continuity, that allows to understand it (in particular how irreversible processes follow from conservative forces). He equally sets to himself that the theory should ensue only from the two first principles of thermodynamics (target constraints). The drama for Planck was that he had to abandon the continuity criterion in order to fulfill adequacy to the experimental data and sufficiency of the two first principles of thermodynamics (a criterion for which he was ready "to sacrifice every one of (his) previous convictions about physical laws" (Planck, 1931)). However, in this case, we see how potent were the set of constraints deemed to be satisfied, and how they even forced a completely new vision of the physical world, not wished for at first, whereby *quantum* physics followed.

Adequacy to the world and target constraints as defined above are therefore enough to specify target conceptual domains *in a normative way*. We are however interested in more, in that we look for a prescriptive way of building new conceptual domains, and even more precisely, we look for cognitive models of learning that could be amenable to computer simulations.

In this framework, learning a new conceptual domain implies both the acquisition of new conceptual primitives with which it then becomes possible to segment the world differently, and the mastering of new control knowledge allowing to efficiently construct models of the environment. In this respect, there are three relevant subproblems :

- *Pattern matching* : how is the world to be matched against new, ill-known conceptual primitives ?
- *Models building* : how ill-known conceptual primitives are combined in order to build models of the situation, and, in addition, how these models are articulated, or articulations, between interpretation domains, in particular the one in gestation and the currently more operational ones ?
- *Conceptualization and theory building* : how a new conceptual domain comes to existence in interaction, both positive and negative, with existing ones ?

These three issues may guide the analysis of human subjects understanding the world (cf. Ohlson, 1996), be they students, scientists or else. They are also key points to be answered in

¹ We will use both terms interchangeably in this paper, on the ground that we are ultimately interested in giving an artificial intelligence account of the discussed inference mechanisms, and that in AI one's interpretation is represented as a model (some symbolic construction or system) standing for the situation at hand.

order to provide a machine intelligence account of the learning of new ways of interpreting the world.

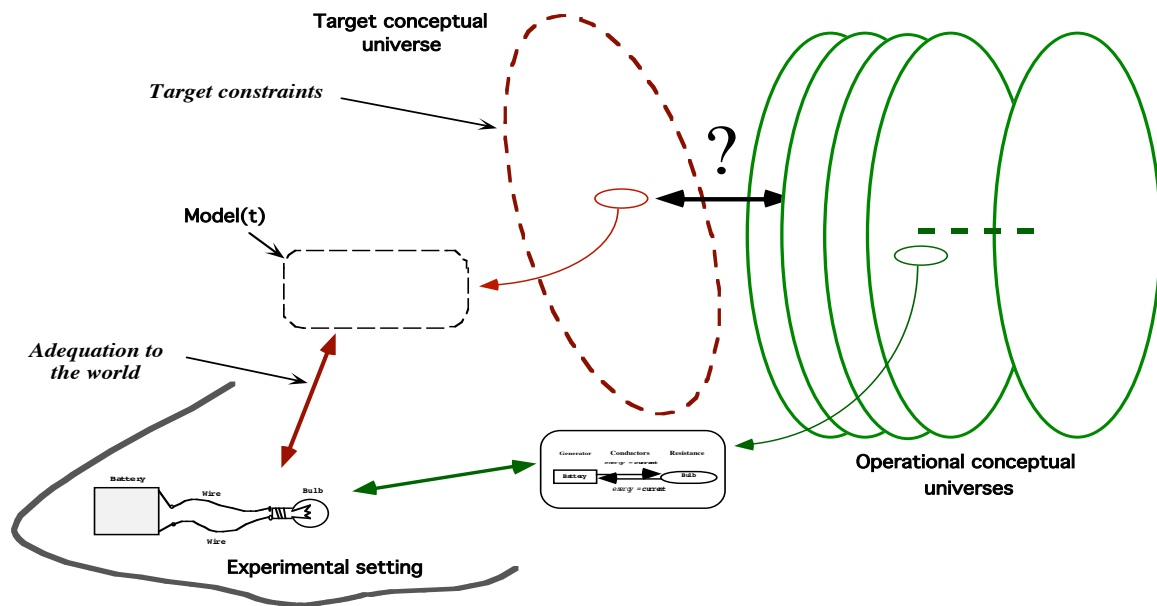


Figure 1. A view on the problem of learning a new conceptual domain.

Figure 1 depicts the main aspects of the overall problem as we see it. The cognitive agent tries to interpret its environment (for instance an experimental setting consisting in a battery connected to a bulb through two wires) in terms of a conceptual domain, in such a way as to satisfy both the adequacy to the world criterion (for instance correct predictions) and the target constraints (for instance, and roughly, to provide an interpretation in terms of energy transfers). At time t , the agent is not able to do that, but because per force he/she knows already a lot about the world, there are known (both operational and activable) conceptual domains that allow him/her to build models of the world, however unsatisfactory they may be in regards to the criteria. This implies that connections be established between an existing and compelling perception of the world, the associated knowledge and a theoretic model under construction, all forms of knowledge that are quite disparate. Furthermore, the new theory, because it is described using known terms (such as "heat" or "reservoir"), and the knowledge about objects and events involved in the perception of the situation, are related to other interpretative domains (such as knowledge pertaining to electricity). As a result, the discovery and acquisition of a new conceptual domain not only requires that connections be set up between the knowledge directly associated with the perception of the situation and the theory, but also that this be done in the context of other, diverse and more or less related, knowledge domains (Tiberghien, 1994, 1996). All of this characterizes nicely the challenge facing both the student and the scientist, as well as many other actors in situations of conceptual learning.

2.2 Experiments in the physics of energy teaching

In order to study learning of new conceptual domains, we set up interpretation tasks in terms of a "new theory". The idea was to force cognitive agents to learn a new way to interpret the world, and to study how they tend to do it. Concentrating on the three subproblems underlined above : segmenting, model building and reconceptualization, we specially focused our attention on pattern matching attempts between the "world" and the new theory, models built in the course

of the task, difficulties encountered, dead-ends, and “repairs” that paved the way to the gradual mastering of a new conceptual system.

Because a scientist or a student trying to come to terms with a set of ill-understood phenomena brings to his/her efforts a whole store of prior knowledge and beliefs, which constitutes a hard to delimit kind of meta control parameter, we organized experiments on human subjects in a way that allowed to control as precisely as possible the prior knowledge they would activate to bear on the problems that were given to them. In this way, it was possible to isolate and analyze the interpretation activity and problem-solving strategies from the activation processes that are responsible for summoning relevant knowledge sources in some context.

More specifically, we performed experiments in physics teaching, and more precisely teaching a qualitative account of the physics of energy taught in high school classes around the age 16-17. The task involved small experimental settings that the students could experiment with, like small electrical circuits with masses and motors and so on, that were to be interpreted in terms of energy transfers and transformations along an “energy chain” starting and ending with an energy reservoir. The students work in pairs². This experiment has been done in several contexts class and in Andrée Tiberghien's laboratory. We video-recorded several pairs of students and entirely transcribed their verbal productions. (For this paper 7 pairs were deeply analysed).

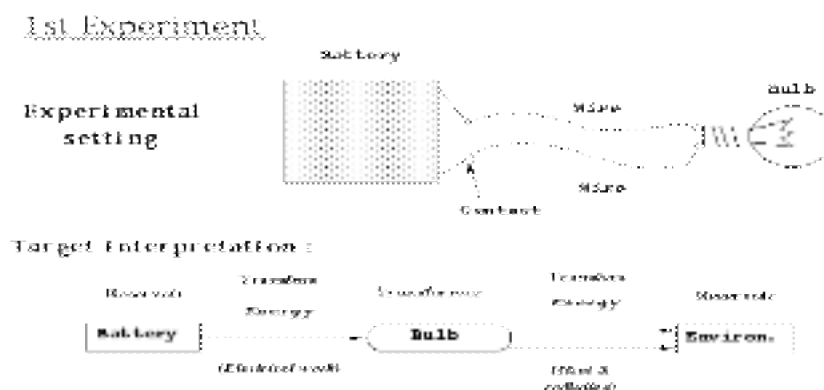


Figure 2. Above : one experimental setting involving a battery connected to a luminous bulb through two wires. Students were to produce an interpretation of this setting in terms of a chain of energy transfers and transformations starting and ending with an energy reservoir. Below : a correct interpretation, called target interpretation.

On one hand, it is important to notice that the interpretation task was not trivial, even in the simplest of the experimental settings shown in figure 2. For instance, there were two wires from the battery to the bulb which satisfied the closed electrical circuit condition, but only one counterpart, standing for the transfer of energy under the form of electrical work, in the target

² In fact the students are given successively three tasks, only the first task is discussed in the paper. In the first task the experimental material is made up of a bulb, two wires, a battery. In the second task the experiment consists of an object hanging on a string which is completely rolled round the axle of a motor (working as a generator). A bulb is connected to the terminals of the motor. When the object is falling, the bulb shines (figure 3). In the third task the experiment consists in a battery connected to an electrical motor. An object is hanging from a string, attached to the axle of the motor, which is completely unrolled at the beginning. A correct solution is given to the students after the first task.

interpretation. Likewise, the students had to discover the environment entity while there was no concrete, tangible, counterpart in the experimental setting.

On the other hand, the task facing the students was easier than the one facing the scientists in that they did not have to “invent” the concepts necessary for the task. They were indeed provided beforehand with a declarative account of the target conceptual domain along with a lexicon of the authorized terms and icons that were to be used in their models of the situation (see figure 3). This is one way we were able to control the knowledge brought to bear by the students. In particular the students were asked to use primitives like *reservoir*, *transformer*, and *transfer*, for which they already had preconceptions (and how to do otherwise, except, may be, by some lengthy and convoluted paraphrase ?). These preconceptions helped them to understand the seed theory, but of course they could also hinder later proper conceptualization.

The seed target domain also defined integrity rules that specified valid models, as, for instance, the “a complete energy chain starts and ends with a reservoir” rule.

Together, the lexical entities used in the definition of the seed conceptual domain and the integrity rules constitute the target constraints for this particular task.

Theory (seed)	Model (seed)
Energy can be characterized by: * its properties : - <i>Storage</i> - <i>Transformation</i> - <i>Transfer</i> - by <i>work</i> : mechanical or electrical - by <i>heat</i> , - by <i>radiation</i> . *a fundamental principle of conservation The energy is conserved whatever the transformations, transfer and forms of storage	* Symbols to be used: r es. for reservoir  for transfer tr . for transformer * Under the constraints: - a complete energy chain starts and ends with a reservoir; - the initial reservoir is different from the final reservoir.

Figure 3. A simplified version of the seed for the target conceptual domain given to the students. The left part presents the conceptual definitions for the target domain . The right part provides the symbols with which to express the model and the syntactic rules that should be satisfied.

One fact that emerged from this study was that out of 7 pairs of students, 6 produced the intermediate model of figure 4 (b) for the battery-bulb setting. They then departed from it to try to find alternatives, better suited models, meantime laboring over concepts like *energy*, *transfers*, and so on. This, in fact, did not strike us as worth of interest at first, so much it appeared to be expected. This intermediate model was after all none other than the classical circular electrical interpretation of the setting. Yet, upon reexamination, we were intrigued by the fact that this model, which acted as a powerful attractor, seemed also pivotal to enable further conceptual elaboration. Did the analysis of the why and how of this particular behavior could lead to a better understanding of the processes at play in the learning of new conceptual domains ? The rest of the paper is an answer to this.

3 The tunnel effect

3.1 Analysis of the experiment

In the experiment described in section 2.2, had the students been experts in the domain of energy chains, they should have produced the model of figure 4 (a). Instead, the vast majority produced at some point the model of figure 4 (b), which clearly looks like model (c) which corresponds to an interpretation of the experimental setting in terms of electricity. However, they were not committed to an electrical interpretation, but were genuinely seeking an interpretation that would satisfy the constraints that were provided to them, that is an interpretation in terms of energy chains (cf. figure 3). How to explain this discrepancy ?

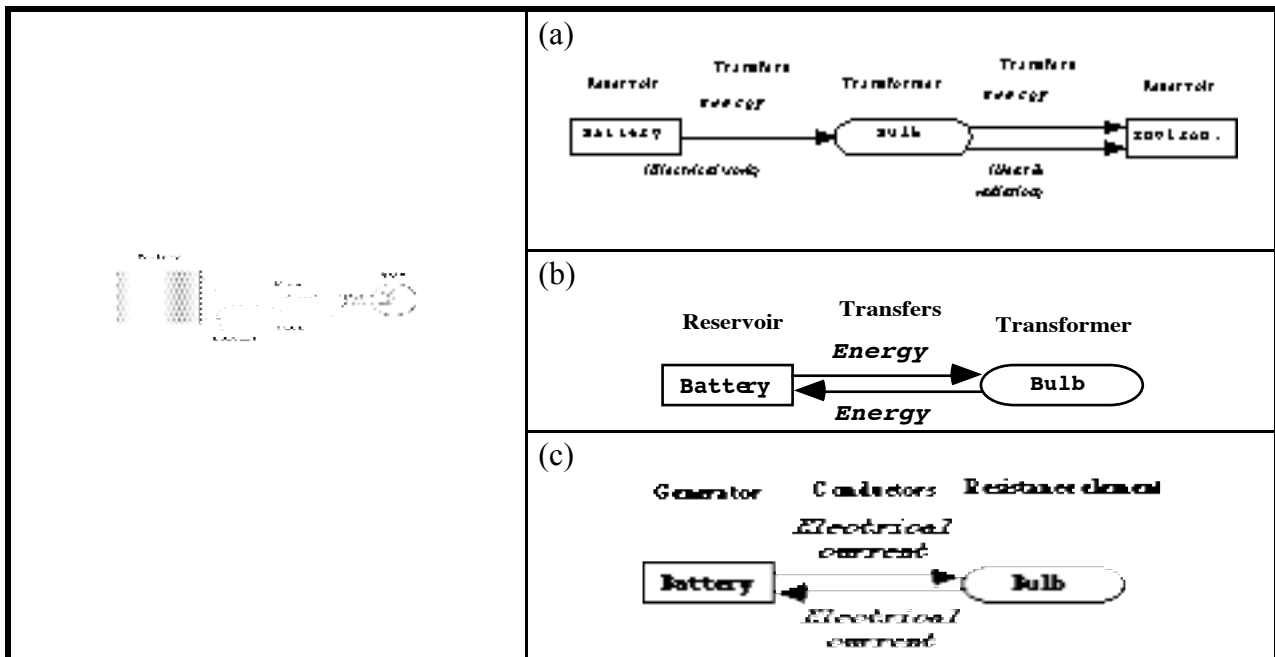


Figure 4. Three interpretations of the experimental setting of the left column.

First of all, it is essential to realize that there is no such thing as an a priori unique, correct and complete representation of the reality. Any description of the world does depend on the subject's current understanding. However, there are descriptions that are very active in some situations and shared by a large set of individuals from the same cultural context. Thus, in our occidental culture of the turn of the third millennium, when someone sees some rather specific shape like in figure 5, he/she cannot escape to see a *battery*, even if on an exploratory mission on Mars.



Figure 5. A shape that we almost necessarily designate by the category "battery".

Hence, depending on the context, there are categories, properties, relations and so on that literally impose themselves in the foreground. They are accordingly the ones that are used when communicating about the situation. Consequently, and quite naturally, the students in our experiments use words like "the battery", "the wires", "the lamp" to point out things in the word. Not only do they use these words to communicate, they also think about the situation using the

associated entities. For instance, suppose you are on an first exploratory mission to Mars, and you see a few steps ahead of you a shape like in figure 5. Not only will you describe it as a battery, but you would seem literally out of your mind if you did not also jump out of surprise and start considering all the implications, including the possibility of the existence of some extraterrestrial life forms knowing of electricity and having set foot on Mars. Referring to some set of perceptions using some word indeed goes far further than merely uttering a designation, it brings with it a whole lot of expectations, constraints on the world, and generally associations to other conceptual structures.

If we then look at the process by which the students built their interpretation of the experimental setting, they start by trying to recognize the matches between the outstanding categories they identify like “the battery”, “the wires” and “the lamp” and the target concepts that have been given to them in the seed theory. They thus match without much trouble the *battery* with a reservoir, and the *lamp* with a transformer³. Without entering into details here, this is rather easy because, on one hand, *batteries* and *reservoirs* share a lot of common properties like being subject to be full or empty, or to play a causal role in many situations, and, on the other hand, *lamps* transforms electricity into light (and heat), and thus is an instance of a *transformer*.

Martin (27) : The reservoir, it will be the battery

Sara (28) : Yeah, yeah.

Lionel (163) : ... the reservoir what is it ? Stores the energy. It is the battery, it is the battery that we put here ...

(The students try to put a name to an arrow)

Fabien (423) : ... what do we write ?

Peggy (424) : It is ... hum ... we write energy, do we ? If not ...

Fabien (425) : Yeah

Peggy (426) : The movement of electrons

Lionel (125) : ... But may be we have to draw the arrows to show where the current goes

Fulvia (126) : But we do not know where it goes

Lionel (127) : From the terminal + to the terminal -

Figure 6. Some instances of dialogs between pairs of students during the task (from Megalakaki & Tiberghien (1995) and Megalakaki (1995)).

Now, when they come to consider the *wires*, it is clear that in the context of this experiment –a physics course and a typical example of an electrical circuit–, the most operational and highly activated interpretation domain is the electricity domain as it has been previously learned in the physics classes. Consequently, and again without entering into details, the *wires* are matched with means for energy transfers, and the *electrical current* is matched with transferred energy. (See figure 6 for instances of exchanges between pairs of students solving the problem). This

³ We use a special font to indicate a conceptual target entity, while we use italics to indicate some entity considered at a notional level, that is outside a specific theory.

electrical counterpart to energy transfers allows to import the naturally inferred (within the electricity interpretation domain) directions of the electrical currents and to make them also the directions for the energy transfers.

It is interesting to see that in this process, compelling interpretations of the world, related to *batteries*, *lamps* and *electrical circuits* have naturally been incorporated into the built model, they have also completely shaped it. Thus, in particular, the commanding electrical interpretation of the setting has led to the model (b) of figure 4 which is akin to the electrical interpretation of model (c).

Thus the overall interpretation, even though it is based on foreign pieces of interpretation and is deeply of an electrical nature, fits most of the syntactical constraints for the satisfaction criteria, and can be mistaken for a valid energy chain interpretation of the setting. This is possible due to the fact that some target entities like *reservoir* or *transformer* match at least partially concrete⁴ entities like *battery* and *lamp*, and thanks to the apparent conformity to the syntactical constraints. This is also due to an apparent synonymy for the students between *current* and *energy*.

But if the interpretation activity stopped there, that would not lead to any further “discovery” by the cognitive agents, nor to the reconceptualization that is part of the learning of a new conceptual domain. There would simply be an electrical interpretation of the phenomenon disguised as an energy chain one. What is interesting is that, except in one case, all the students then embarked in further predictive activity based on the model they just came up with. They thus re-interpreted the model stemming from electrical considerations within the energy interpretation domain ! That is they suspended the underlying *raison d’être* of the model to interpret it within a new conceptual domain. This seemingly instantaneous autonomization of the model and reinterpretation within a new conceptual domain, is reminiscent of the *tunnel effect* in quantum physics whereby a particle can occasionally tunnel through barriers, or, more precisely, escape a potential well to enter another one without having enough energy to overcome the potential barrier between them. In our case, the passage from one interpretation domain to another one should go along with a dismantling of the first interpretation/model before reconstructing a new one in the new domain. Instead, the very same model becomes autonomous from its former source domain and is reinterpreted as such in a new one as if a tunnel had been drilled between the two domains allowing one to go from one to the other unnoticed and inconspicuously. Before analyzing how this is possible, it is interesting to see what happens next.

Some students, “reading” the consequences of their model, predict that energy will come back to the battery, which they know, from their prior background knowledge, is not possible. These students are thus in front of a violation of the adequacy to the physical world criterion. Others (a minority) realize that their model does not satisfy the integrity rule of the seed theory according to which the initial and final reservoirs should be different. In any case, an intense reexamination of the model, its meaning and its justification takes place, leading to a better mastering of the

⁴ “ Concrete entities ” are of course no more concrete than conceptual ones. As Schrödinger (1982) stated, “ chairs and tables are, as much as state vectors (in quantum physics), intellectual constructs deemed to articulate our experience around a small set of invariants. Chairs and tables are theoretical entities to the same extent as state vectors. The only difference being that the theory in which chairs and tables are integrated is the one we had to build since our early youth in order to survive ”. Thus concrete entities, like chairs and tables, refer to class of very familiar objects whereas, for the learner, energy has no direct correspondence in the physical world.

target domain (Cauzinille-Marmèche et al. (1997) provide an analysis of the “repair mechanisms” used by students to adapt their model).

3.2 The tunnel effect mechanism

In its broad lines, the tunnel effect is very simple to define. It involves two main stages :

1. A source (possibly composite) interpretation of the experimental setting *disguises itself under the dresses of a legitimate model in the target interpretation domain*.
2. This model, bare of its underlying justifications, is *then re-interpreted within the target interpretation domain*.

There might follow unforeseen predictive consequences of this model, possibly leading to adaptation and reconceptualisation in the target domain.

As an illustration, going back to the energy chain interpretation task, we can distinguish the two following stages :

1. From a multiplicity of considerations involving *batteries, lamps, electrical wires* (categories that are outstanding in the context of this task and are necessary for the purpose of description and communication among the students) and the provided seed theory with its reservoir, transfer, transformer and energy entities, the students built a model of the setting that can appear to be a valid model within the target interpretation domain.

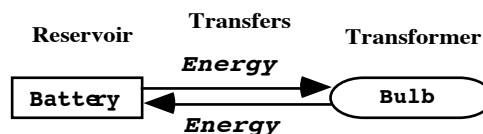


Figure 7. The intermediate model for the battery-bulb experimental setting.

In particular, remember that under the label “energy” in this model, there is the association with *electrical current* that has allowed to endow these energy transfers with directions.

2. In a second stage, this model is then re-interpreted within the target domain of energy chains. Consequently, it is predicted that energy will flow back to the *battery* which is also the final reservoir of energy. And this prediction does not satisfy the students who, at this time, hold an energy point of view. This leads the students to challenge the current model.

Note that there really is reinterpretation taking place here. If not, that is if the label “energy” was still attached to the underlying meaning of electrical current, there would not be any problem with the energy (electrical current) going back to the battery. Indeed, one pair of students in our experiment was contented with its intermediate model and never reached the re-interpretation stage.

Note also how unlikely it is that the model of figure 7 would have been produced directly within the target interpretation domain. There is no a priori reason for the arrows going forth and back, and this arrangement violates the integrity rule that says that the initial reservoir should be different from the final one.

Section 3.4 provides another illustration of tunnel effect in the case of the discovery of thermodynamics by Carnot, Joule, Thomson and Clausius.

To sum up somewhat boldly what has been said, a tunnel effect can occur each time a model (something that is expressed with symbols and has enough duration to be re-examined) becomes autonomous from its initial justifications and is liable to be re-interpreted in a new interpretation domain.

Now, there are several questions that need to be considered in order to have a more operational specification of tunnel effects :

- How some (possibly composite) interpretation can disguise itself under a valid model within a target interpretation domain ?
- What can be imported, and how, from the source interpretation to the re-interpretation domain once only the expressed model remains without its underlying *raison d'être* ?

We study each of these questions in turn.

1. *How a target interpretation domain can a priori, before being completely set up, specify what a valid model should look like ?*

Our experiment in physics teaching is a special case where the target interpretation domain is provided ahead of time and outside any “project ” from the students. Hence, the seed theory can be completely a priori specified. Here, it imposes various constraints on any candidate model of the world by imposing a formalism and some qualitative integrity rules (e.g. “an energy chain should start and end with a reservoir ”). In actual scientific discovery processes, the specifications for target models are set up a priori from preconceptions about the target domain. For instance, Maxwell in its search for a theory for electromagnetism was looking for models of the phenomena with continuous interactions rather than Newtonian like actions at a distance, and thus was looking for a formalism relying on the differential calculus. As we stressed already, Planck was adamant that his models would ultimately rest only on the two first principles of thermodynamics. Carnot, in thermodynamics, was looking for cyclic and reversible models of the steam engines in order to be able to establish their maximal efficiency.

In any case, even though the target conceptual domain is per force not completely defined a priori, there always are minimal criteria that specify how a valid model should look like. These criteria may be erroneous, of a temporary nature, partial, but they always exist as a projection of what is believed to be fundamental in the target domain. In this way, many models, that may turn up to be unsatisfactory in face of measured phenomena or because of inconsistencies, may appear at first as legitimate candidates. Any model, whatever its underlying justification, that thus satisfies these general requirements is open to be interpreted within the target domain. And any model that was built at least partially outside the target interpretation domain, and that appears nonetheless legitimate in this domain can be said to be disguised.

2. *What can be imported, and how, from the source interpretation to the re-interpretation domain ?*

Once again, it is helpful to look at the energy chain task. At one point, the students are ready to associate *electrical current*, that they literally “see” in the experimental setting, with the target entity energy. They do this presumably because, at the notional level where *current* and *energy* share many properties like being fluid, circulating, being agents for causality, and so on. Once this association has been approved, then everywhere in the model that electrical current would appear, it is replaced with the label energy. But what is essential is that, at the same time, everywhere some properties of energy are needed (like its transfer direction) in the model, this is the properties associated with electrical current that are imported. And these are not thought

upon and pondered, but on the contrary, they are smuggled in without further immediate checking. Hence the circular nature of the model of figure 7.

It is important to realize that this phenomenon, which is central in what we call the tunnel effect in cognition, is ordinary. It happened when Carnot was equating the “caloric” with heat, and thereby introducing —smuggling in— its conservative property. It happened to Maxwell when he equated the ether (incompressible fluid) with a model for electromagnetic interactions, smuggling in the seeds for the difficulties faced in physics until Einstein’s special relativity theory got rid of them (and of most of the smuggled in properties of ether). It happens all the time, and it happens unconsciously. This smuggling might turn out to be genial when it brings with it unexpected solutions to outstanding problems. It might also hinder further solution. We discuss these aspects in section 3.4 below.

To sum up, each time entities from two different interpretation domains are matched, they can potentially bring with them in these associations further attached properties that are new to the other entity. And this can happen in both ways. For instance, we noted that energy transfers found themselves naturally endowed with directions as soon as energy was matched with electrical current. Likewise, in another task not presented here, one student matched reservoir with a weighting object, then to show that the weight could be filled up (!) by being lifted. An example of a property not to be found originally in the notion of weight (source domain), but really brought by the contextual match with reservoir (target domain).

Each time a new conceptual domain is learned (either by being taught or by discovery), it is unavoidable that, at the start, it is related and articulated to the current (situated or contextual) operational interpretation domains. (At the start, student cannot speak about “the reservoir” in the setting. They must use designations like “the battery” that are operational to them at this time). Because of this, when learners build models of the environment in the new target conceptual domain, they necessarily do this by matching entities and structures from the operational domains to the target one. What we argue here is that, most of the time, by doing this, there are hidden properties that are smuggled in these matching operations. These properties shape the model built in unchecked ways. This is only when re-interpretation occurs entirely within the target domain in construction that these hidden aspects may reveal themselves, thus bringing out unforeseen consequences in the target interpretation. This phenomenon that we call tunnel effect is therefore responsible for transferring information from the source domain(s) to the target one. Section 4 below discusses this information transfer property in comparison with analogical reasoning, the most well-known inference mechanism for transferring information between domains.

3.3 Tunnel effect as a way to decompose problem solving

One can see tunnel effect as a way to ease problem-solving in an ill-mastered conceptual domain. For instance, no students were able to solve directly the first energy chain task. The problem was simply too hard for them. On the other hand, 12 out of 14 produced the intermediate model of figure 7, which is arguably an electric model of the setting disguised as a legitimate model in the energy domain. If it is difficult for an agent to solve directly a problem in an ill-mastered domain, it might be easier to disguise an interpretation stemming from well-known domains into a legitimate candidate model in the target domain. The question then is of course that of seeing if

that step (a kind of forgery, except it may be unconscious resulting from automatic inferencing in the source domain(s)) helps or hinders further resolution of the problem.

It is difficult to answer this question in general, except that some version of the now famous *no-free lunch theorem* known in Machine Learning and Optimization Theory is likely to apply and state that, overall, tunnel effect must equally ease and hinder problem-solving in new domains depending on the context.

However, there are reasons to think that tunnel effect may be a powerful help in problem-solving in some cases. Figure 8 suggests why. Thanks to tunnel effect, there are apparently more solutions to the interpretation problem, and hence more opportunities to find one of them. The problem then, if a fallacious solution has been found, is to be able to find a way towards the correct solution. We show in section 3.4 that this may be facilitated by the focus naturally provided by the processes underlying tunnel effect.

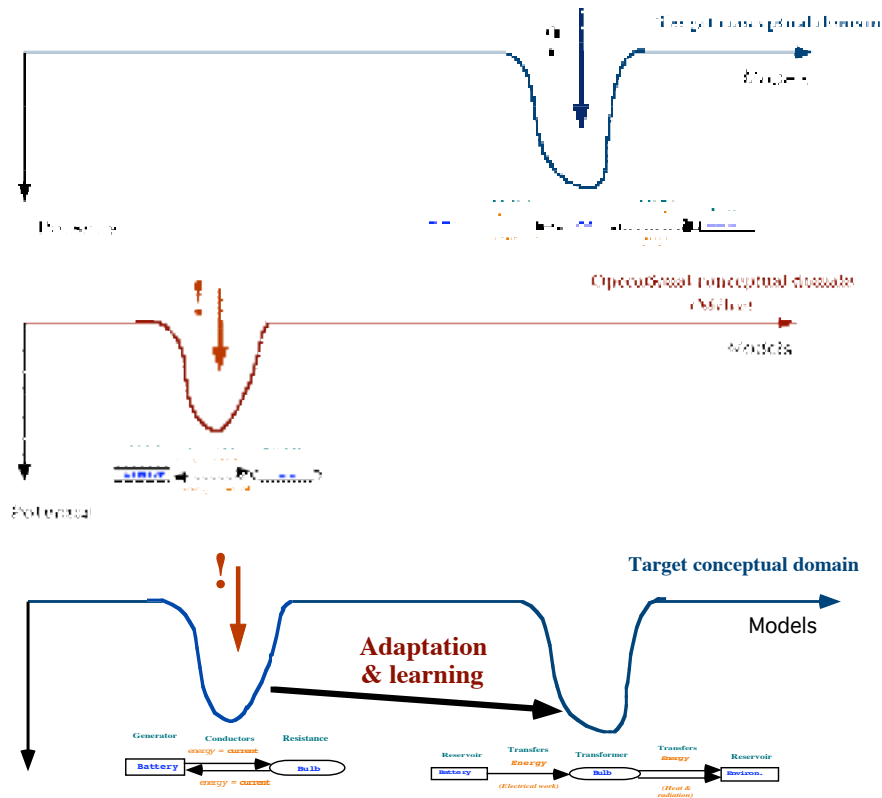


Figure 8. A decomposition process facilitating problem-solving in an ill-mastered conceptual domain. In each schema, the horizontal axis stands for the space of potential models and the vertical axis stands for the quality of the model with regards to the world. Of course, depending on the interpretation domain (for instance centered on electricity or on energy exchanges), the same models may have different degrees of quality.

3.4 How tunnel effect activates further adaptation and conceptual learning

Two cases must be examined with respect to the opportunities for learning opened when a model has been obtained using tunnel effect :

1. The model obtained remains valid even after being re-interpreted in the target domain under construction.

2. The model turns out to be erroneous either when confronted with the world or because internal inconsistencies are discovered within the target interpretation domain.

We study these two cases in turn.

1. The model remains valid.

This is what happened during the construction of thermodynamics by Carnot, Clapeyron, Thomson, Joule, Clausius and others (Longair, 1984; Science & Vie, 1994). Carnot, influenced by the theory of the *caloric* (an imponderable fluid with the property of being conserved and which he equated to heat) and by his father's work on the calculation of the efficiency of water mills, devised a cyclic and reversible model describing an ideal steam engine. Thanks to this model, he was able to demonstrate that there exists a maximal efficiency for steam engines, and that it depends on the difference of temperature between the hot source of heat (caloric) and the cold one. Later on, through a series of very meticulous experiments, Joule was able to show that heat was not a conservative quantity and was exchangeable with work. However, it turned out that Carnot's model was in fact neutral with respect to the caloric hypothesis and when re-interpreted in the context of the new theory about heat and work, still remained a very helpful tool for thought experiments, one which eventually led to the discovery by Clausius of a special state function called entropy.

We have here one instance of a model obtained through tunnel effect (its cyclic and reversible nature was deeply a result of the belief in the caloric theory even though this was never explicitly expressed by Carnot) which is still valid once the interpretation domain changes. The model by itself cannot therefore act as a trigger for re-evaluation of the target domain, and other symptoms must show. However, because it remains valid, it can help shape the new conceptual system and serve as a test bed for it, potentially through thought experiments as this was the case for Carnot's model in thermodynamics.

2. The model turns out to be erroneous when re-interpreted.

In our energy chain experiments, this happened either when students realized that the model implied that the energy was flowing back to the battery (which they knew was incorrect), or when they discovered an inconsistency with the target integrity rule stating that the initial energy reservoir should be different from the final one.

The natural question is then why is the model wrong in the investigated aspect ? A re-examination of the path that led to this conclusion in the model can then point towards one of two causes. First, the associations made between entities from the target domain and the source one(s) could be erroneous. For instance, many students question the association they made between electrical current and energy or between the wires and the transfers. This can lead to a differentiation process whereby the target entities gain autonomy with respect to the source ones. Second, the automatic inferencing process that determined the problematic aspect of the model can be disclosed and limitations for its range been set. This is what happened when some students realized that the circular nature of the electrical current did not carry to the energy entity. This inference was henceforth stopped when building a model.

This short discussion convincingly shows in our opinion that tunnel effects, not only help finding models, even erroneous ones, but that they also provide guidelines for further re-examination and reconceptualisation when needed. This is however an issue that deserves much further work.

4 The tunnel effect vs. analogical reasoning

Very few inference mechanisms have been proposed that deal with the transfer of information between different conceptual domains. *Analogical reasoning* is one of them —the most famous—, *blending* is another one (Fauconnier & Turner, 1998), and, we submit, *tunnel effect* is a contender too. A full comparative study of the three of them would be more than interesting, but is beyond the scope of this paper. However, we believe that a comparison with analogical reasoning might help to enlighten some characteristics of the tunnel effect as an inferencing mechanism. We will concentrate in each case on the conditions for a transfer between interpretation domains to occur, and on the information content that is transferred.

According to the dominant view on analogy (e.g. (Falkenheimer et al., 1989; Greiner, 1988)), **analogical reasoning** involves the interpretation of two cases, —called the source case for the supposedly well-known one, and the target case for the one to be completed—, that may be interpreted within two different interpretation domains (e.g. the solar system as a source case and the supposedly ill-understood atom system as a target one). Each case is supposed to be represented as a graph of relations and nodes standing for primitive concepts. Analogical reasoning implies then that a best partial match be found between the two graphs, and, in a second step, that the part of the graph representing the source case with no counterpart in the target case representation be copied, translated and added to the target representation in order to fill the missing part. Many questions arise as to the principles that should govern both the matching operation, the translation and the transfer, not to speak about subsequent verification and adaptation. Deep concerns have also been expressed about the interpretation process of the two cases during analogy and the ensuing representation of the cases (e.g. (Hosftädter, 1995; Mitchell, 1993)). It is important to note that both domains —the source and target— must be sufficiently well understood in order that the respective conceptual primitives be identified, put in hierarchy and potentially matched. This view of analogical reasoning thus prevents the consideration of a target domain that would be in gestation and of which conceptual primitives would be very uncertain.

If we consider then the analogical inferencing mechanism as a kind of black box with inputs and outputs, the *inputs* consist in the source and target conceptual domains (the conceptual primitives and their relationships (including the said over-important hierarchies) and in the two cases (be they already represented as some would pretend is realistic or be they interpreted in the context of the analogy as others would insist is unavoidable). The *black box* then searches for one satisfying matching between the two cases (given as rigid representations or not) and computes the completion of the target case representation. The *output or information gained* in the operation consists therefore in the added features and properties of the target case.

In contrast, **tunnel effect** only involves the interpretation of a *single situation or case* (e.g. an experimental setting or a set of phenomena). The *input* of the tunnel effect black box consists in the operational source interpretation domain(s), the target criteria that specifies the target interpretation domain (including preconceptions about some target entities, their properties and relationships), and the case (situation or set of phenomena) to be interpreted and understood in the target interpretation domain (e.g. the battery-lamp experiment to be interpreted in terms of energy exchanges, the electromagnetic interactions as measured in Faraday's experiments in terms of a theory in germ in Maxwell's head, or the steam engines in terms of heat and work and other related variables in the nascent thermodynamics). The *black box* then searches for a model of the case satisfying the target criteria. Because most target entities are not yet operational and

interpretable directly in the world, they have to be translated in terms of the more operational interpretation domains given as inputs. In this translation process, submitted to the target criteria, and during model building, some aspects of the model may be automatically filled up through automatic inferencing within the source domain(s) (as is the case when the arrows for transfers are automatically specified when it is decided to translate energy transfer from the notion of electrical current). The *output or information gained* in the operation consists in the unexpected (because not planned) consequences of the model when interpreted within the target interpretation domain, or in the experimental setting if some target entities are already partially interpretable in the world (as is the case for "energy" for 16-17 years old students).

Analogy	Tunnel effect
• Two experimental settings or situations that are posited as analogs to each other • Interpretation takes place both in the source domain and in the target domain (there are two situations to be interpreted). • Relies heavily on comparisons : • Implies complex pattern matching between the two case representations • Tightly associated with the notion of similarity <i>between</i> structures. One problem is to explain how this similarity is computed • There is transfer by matching, alignment and completion from the source to the target • <i>New information</i> is produced through the completion of the target case representation • Does not explain how the source is chosen • Learning is supposed to arise as : - learning of indexing scheme - generalization and abstraction from analog cases - not really new conceptualization, except by generalization	• One experimental setting or situation only • Interpretation takes place in the source domain subject to the target constraints and adequacy to the world criterion. • No comparison is involved, only interpretation • Involves associations at the notional level between target entities and source ones • Associated with confusion at the notional level. No notion of similarity <i>between</i> constructs • There is transfer by reinterpretation of the model of which some aspects have been automatically filled-in within the source interpretation domain(s). The built model gains autonomy and is reinterpreted in the target domain • <i>New information</i> is produced through automatic completion of the model within the source domain • The source domain(s) is(are) the most operational for interpretation in the current situation • Learning : - Reconceptualization focuses on associated entities that led to inconsistencies in order to differentiate them - Progressive operationalisation of the new conceptual domain - Articulation with primitive perceptions about the world and with the source conceptual domain

Table 1. A summary of the main features of analogical reasoning versus features of tunnel effect.

In both analogical reasoning and tunnel effect, the detection of discrepancies between the resulting model and the world or of other inconsistencies opens opportunities for learning. The difference lies in the fact that tunnel effect is intrinsically intended towards the process of building the domain interpretation domain (through the setting up of connections between this domain, the operational ones in the context and the world) whereas analogical reasoning is oriented towards the completion of some specific case with the help of another 'similar' one.

While failed analogies may lead to reconceptualisation in the target interpretation domain, this is much less direct than the learning that may occur when a tunnel effect has produced an unfit model of the world in the interpretation domain.

5 Conclusions

This paper takes seriously the idea that cognition may imply the existence (and coexistence) of several different interpretation universes, and that a specially important type of learning consists in acquiring new ways of interpreting the world or some aspects of it. In our study we focused on the passage from the currently operational interpretation domain(s) to a new target one when the attention of the cognitive agent is driven towards the interpretation and understanding of some phenomenon or set of phenomena.

In studying the type of conceptual learning at play when students are learning a new conceptual domain or when scientists are struggling to find new ways to account for the world, we discovered the pivotal role of intermediate expressed models.

Indeed, when a new interpretation domain is learnt (i.e. new segmenting of the world and new inference rules), the new concepts and new rules are not yet settled nor directly interpretable in the world (think about the first time you heard of tensor calculus or of electrons). They have to be linked with known entities. Therefore, when a model is built in terms of target entities, it in fact refers to the world mostly through entities and relations belonging to the currently operational domain(s). Aspects of this model might thus be filled in thanks to automatic (and unchecked) inferences within the source domain(s). This is the basis for the tunnel effect. These added features, expressed in the model, when re-interpreted within the target domain may bring out unforeseen consequences.

Tunnel effect is thus a special inference mechanism at play when models are built at the intersection (but not quite in fact) of some operational interpretation domain(s) —with its/their automatic inferencing capability— and a new ill-known one. Tunnel effect is ubiquitous, mostly unconscious and central in the learning of new conceptual domains. It has so far, to the best of our knowledge, not been described and studied.

Tunnel effect eases the construction of models by providing inference mechanisms from the source domain(s) that make up for the as yet non-existent inference mechanisms of the target domain. In so doing, erroneous models might be obtained. These intermediate models can help or hinder reaching a later, more adapted, model. Even though we think we have strong arguments to the effect that tunnel effect can be a powerful guide for further reconceptualisation (see section 3.4), this is still a matter for research, specially in view to the fact that, in case favorable conditions could be identified, one could envision using well-guided tunnel effects to ease the teaching of scientific domains.

References

- Cauzinille-Marmèche, E., Collet, G., Cornuéjols, A. & Tiberghien, A., 1997, Co-adaptation of students' knowledge domains when interpreting a physical situation in terms of a new theory. In Proc. of the 2nd European Conf. on Cognitive Sciences (ECCS'97), Manchester, April 1997, pp.107-112.
- Falkenhainer, B., Forbus, K.D. & Gentner, D., 1989, Structure-mapping engine, *Artificial Intelligence*, 41, 1-63, 1989.

- Fauconnier, G. & Turner, M., 1998, Conceptual integration networks. *Cognitive Science*, vol.22 (2):133-187.
- Greiner, R., 1988, Learning by understanding analogies. *Artificial Intelligence* 35:81-125.
- Hofstadter, D., 1995, *Fluid Concepts and Creative Analogies*, Basic Books.
- Holton, G., 1973, *Thematic Origins of scientific thought. Kepler to Einstein*. Harvard University Press, 1973.
- Longair, M.S., 1984, *Theoretical Concepts in Physics*. Cambridge University Press, 1984.
- Megalakaki, O. & Tiberghien, A., 1995, Corpus de dialogues de trois groupe d'élèves résolvant trois problèmes mettant en jeu une activité de modélisation, CNRS-COAST Research Report, no. CR-10/95.
- Megalakaki, O., 1995, Expérience MARIANNE : corpus de dialogues de trois groupes d'élèves résolvant une séquence de problèmes mettant en jeu une activité de modélisation, CNRS-COAST Research Report, no. CR-11/95.
- Mitchell, M., 1993, *Analogy-Making as Perception*, MIT Press.
- Nersessian, N., 1992, How do scientists think? Capturing the dynamics of conceptual changes in science in Giere, R. (Ed.) *The Minnesota Studies in the Philosophy of science*, vol.XV, University of Minnesota Press, Minneapolis, pp.3-44, 1992.
- Ohlsson, S., 1996, Learning to do and learning to understand. A lesson and a challenge for cognitive modeling. In P. Reiman & H. Spada (Eds), *Learning in Humans and Machines*, pp.37-62, Oxford, Pergamon.
- Planck, M. 1931, Letter from M. Planck to R.W. Wood. (See Hermann, A., 1971, *The Genesis of Quantum Theory (1899-1913)*, MIT Press, Cambridge Mass., pp.23-24).
- Schrödinger, E., 1982, *Ma Conception du Monde. (My vision of the world)*. Mercure de France, 1982, p.30.
- Science & Vie, 1994, Special issue on Carnot and the discovery of thermodynamics (in French).
- Tiberghien, A., 1994, Modelling as a basis for analysing teaching-learning situations, *Learning and Instruction*, 4(1):71-87.
- Tiberghien, A., 1996, Construction of prototypical situations in teaching the concept of energy. In G. Welford, J. Osborne, P. Scott (Eds) *Research in Science Education in Europe*. London: Falmer Press. p. 100-114.