Abduction as a Rational Means to Creativity. Unexpressed Knowledge and Scientific Discovery

Lorenzo Magnani[1] and **Matteo Piazza**[2] and **Riccardo Dossena**[3]

Abstract. Science is one of the most creative forms of human reasoning. It has been stressed that scientific discovery is a mysterious process, involving irrationality and any kind of unexplainable feelings and emotions. By rejecting that point of view, the recent epistemological and cognitive studies concentrate on the concept of *abduction*. Usually these studies pay attention to those *theoretical* and "internal" aspects of abduction which describe the process in computational terms, and which study it by means of artificial simulations and/or abstract cognitive models. A neglected issue, worth of a deepest investigation inside artificial intelligence (AI), is that "creation" is often a complex cognitive task involving transformations in the state of the cognitive system, from a rough "material" experience of the world to a more sophisticated and structured one. Concrete manipulations of external world is a fundamental passage in the process of knowledge acquisition and hypotheses generation: through a process of *manipulative abduction* it is possible to build prostheses for human minds, by interacting with external objects in a constructive way, and so by creating *implicit* knowledge through doing. This kind of embodied and unexpressed knowledge holds a key role in the subsequent scientific comprehension and discovery. This paper aims to illustrate the close relationship between external representations and creative processes in science, also from the point of view of the so-called unexpressed knowledge.

1 INTRODUCTION

Science is one of the most explicitly constructed, abstract, and creative forms of human knowledge. The problem of scientific discovery and change has been central inside philosophy of science since the mid-1960s, when some critics challenged the traditional picture of conceptual change, offered by logical positivism, as a continuous and cumulative process justifiable by modern formal logic.

Kuhn, by means of his perspective in terms of "paradigms", claimed that scientific change is assimilable to some sort of irrational and not analyzable *Gestalt-switch*, best understood as a conceptual *revolution* involving the overthrow and the replacement of the reigning concepts with others *incommensurable* (see [13]).

In the twentieth century Kuhnian ideas brought philosophers of science to distinguish between a *logic of discovery* and a *logic of justification*, and to the direct conclusion that a logic of discovery, and then a *rational* model of discovery, cannot exist. This was coherent with a cataclysmic and irrational interpretation of the scientific enterprise, justified by the apparent dramatic and discontinuous nature of scientific change. Today researchers have by and large abandoned this attitude by concentrating on the concept of *abduction* pointed out by C.S. Peirce as a fundamental mechanism by which it is possible to account for the introduction of new explanatory hypotheses in science.

Abduction is the process of *inferring* certain facts and/or laws and hypotheses that render some sentences plausible, that *explain* or *discover* some (eventually new) phenomenon or observation; it is the process of reasoning in which explanatory hypotheses are formed and evaluated. There are two main epistemological meanings of the word abduction [15]: 1) abduction that only generates "plausible" hypotheses ("selective" or "creative") and 2) abduction considered as inference "to the best explanation", which also evaluates hypotheses. To illustrate from the field of medical knowledge, the discovery of a new disease and the manifestations it causes can be considered as the result of a creative abductive inference. Therefore, "creative" abduction deals with the whole field of the growth of scientific knowledge. This is irrelevant in medical diagnosis where instead the task is to "select" from an encyclopedia of pre-stored diagnostic entities. We can call both inferences ampliative, selective and creative, because in both cases the reasoning involved amplifies, or goes beyond, the information incorporated in the premises.

Theoretical abduction[4] certainly illustrates much of what is important in creative abductive reasoning, in humans and in computational programs, but fails to account for many cases of explanations occurring in science when the exploitation of environment is crucial. It fails to account for those cases in which there is a kind of "discovering through doing", cases in which new and still unexpressed information is codified by means of manipulations of some external objects (*epistemic mediators*). The concept of *manipulative abduction*[5] captures a large part of scientists thinking where the role of action is central, and where the features of this action are implicit and hard to be elicited: action can provide otherwise unavailable information that enables the agent to solve problems by starting and by performing a suitable abductive process of generation or selection of hypotheses.

[1] Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy, and Georgia Institute of Technology, Atlanta, USA, email: lmagnani@cc.gatech.edu

[2] Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy, email: mpiazza@pure-reason.unipv.it

[3] Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy, email: rdossena@pure-reason.unipv.it

[4] Magnani [15] introduces the concept of theoretical abduction. He maintains that there are two kinds of theoretical abduction, "sentential", related to logic and to verbal/symbolic inferences, and "model-based", related to the exploitation of models such as diagrams, pictures, etc, cf. below in this paper.

[5] Manipulative abduction and epistemic mediators are introduced and illustrated in [15].

2 ABDUCTION AND EXTERNAL WORLD

2.1 The "internal" side of creative reasoning

Throughout his career it was C.S. Peirce that defended the thesis that, besides deduction and induction[6], there is a third mode of inference that constitutes the only method for really improving scientific knowledge, which he called *abduction*. Science improves and grows continuously, but this continuous enrichment cannot be due to deduction, nor to induction: deduction does not produce any new idea, whereas induction produces too simple ideas. New ideas in science are due to *abduction*, a particular kind of non-deductive[7] inference that involves the generation and evaluation of explanatory hypotheses.

All we can expect of our "selective" abduction, is that it tends to produce hypotheses for further examination that have some chance of turning out to be the best explanation. Selective abduction will always produce hypotheses that give at least a partial explanation and therefore have a small amount of initial plausibility. In the syllogistic view advocated by Peirce (see below) concerning abduction as inference to the best explanation one might require that the final chosen explanation be the most "plausible". Since the time of John Stuart Mill, the name given to all kinds of non deductive reasoning has been induction, considered as an aggregate of many methods for discovering causal relationships. Consequently induction in its widest sense is an ampliative process of the generalization of knowledge. Peirce distinguished various types of induction: a common feature of all kinds of induction is the ability to compare individual statements: using induction it is possible to synthesize individual statements into general laws – inductive generalizations – in a defeasible way, but it is also possible to confirm or discount hypotheses.

Following Peirce, we are clearly referring here to the latter type of induction: abduction creates or selects hypotheses; from these hypotheses consequences are derived by deduction that are compared with the available data by induction. This perspective on hypothesis testing in terms of induction is also known in philosophy of science as the "hypothetico-deductive method" [8] and is related to the idea of confirmation of scientific hypotheses, predominant in neopositivistic philosophy but also present in the falsificationist tradition [29].

Deduction is an inference that refers to a logical implication. Deduction may be distinguished from abduction and induction on the grounds that only in deduction is the truth of the conclusion of the inference guaranteed by the truth of the premises on which it is based. Deduction refers to the so-called non-defeasible arguments. It should be clear that, on the contrary, when we say that the premises of an argument provide partial support for the conclusion, we mean that if the premises were true, they would give us good reasons – but not conclusive reasons – to accept the conclusion. That is to say, although the premises, if true, provide some evidence to support the conclusion, the conclusion may still be false (arguments of this type are called inductive, or abductive, arguments). All these distinctions need to be exemplified. To describe how the three inferences operate, it is useful to start with a very simple example dealing with diagnostic reasoning and illustrated (as Peirce initially did), in syllogistic terms:

1. If a patient is affected by a pneumonia, his/her level of white blood cells is increased.
2. John is affected by a pneumonia.
3. John's level of white blood cells is increased.

(This syllogism is known as Barbara).
In this case the inference rule corresponds to

$$\frac{\begin{array}{l}\varphi \rightarrow \psi \\ \psi\end{array}}{\varphi}$$

By deduction we can infer (3) from (1) and (2). Two other syllogisms can be obtained from Barbara if we exchange the conclusion (or Result, in Peircian terms) with either the major premise (the Rule) or the minor premise (the Case): by induction we can go from a finite set of facts, like (2) and (3), to a universally quantified generalization – also called categorical inductive generalization, like the piece of hematologic knowledge represented by (1) (in this case we meet induction as the ability to generate simple laws, constrasted to induction as a way to confirm or discard hypotheses, cf. above). Starting from knowing – selecting – (1) and "observing" (3) we can infer (2) by performing a selective abduction. The abductive inference rule corresponds to the well-known fallacy called affirming the consequent (simplified to the propositional case)

It is useful to give another example, describing an inference very similar to the previous one:

1. If a patient is affected by a beta-thalassemia, his/her level of hemoglobin A2 is increased.
2. John is affected by a beta-thalassemia.
3. John's level of hemoglobin A2 is increased.

Such an inference is valid, that is not affected by uncertainty, since the manifestation (3) is pathognomonic for beta-thalassemia (as expressed by the biconditional in $\varphi \leftrightarrow \psi$). This is a special case, where there is not abduction because there is not "selection", in general clinicians very often have to deal with manifestations which can be explained by different diagnostic hypotheses: in this case the inference rule corresponds to

$$\frac{\begin{array}{l}\varphi \leftrightarrow \psi \\ \psi\end{array}}{\varphi}$$

2.1.1 *Sentential abduction*

Many attempts have been made to model abduction by developing some formal tools in order to illustrate its computational properties and the relationships with the different forms of deductive reasoning (see, for example, [2]). Some of the formal models of abductive reasoning are based on the theory of the *epistemic state* of an agent (see [1]), where the epistemic state of an individual is modeled as a consistent set of beliefs that can change by expansion and contraction (*belief revision framework*).

Deductive models of abduction may be characterized as follows. An explanation for β relative to background theory T will be any α that, together with T, entails β (normally with the additional condition that $\alpha \cup T$ be consistent). Such theories are usually generalized in many directions: first of all by showing that explanations entail their conclusions only in a *defeasible* way (there are many potential explanations), so joining the whole area of the so-called nonmonotonic logic or of the probabilistic treatments; second, trying to show

[6] Peirce clearly contrasted abduction to induction and deduction, by using the famous syllogistic model. More details on the differences between abductive and inductive/deductive inferences can be found in [15, 4].

[7] Non-deductive if we use the attribute "deductive" as designated by the classical logic.

how some of the explanations are relatively implausible, elaborating suitable technical tools (for example in terms of modal logic) able to capture the notion of preference among explanations.

The idea of consistency that underlies some of the more recent deductive consistency-based models of selective abduction (diagnostic reasoning) is the following: any inconsistency (anomalous observation) refers to an aberrant behavior that can usually be accounted for by finding some set of components of a system that, if behaving abnormally, will entail or justify the actual observation. The observation is anomalous because it contradicts the expectation that the system involved is working according to specification. This types of deductive models go beyond the mere treatment of selective abduction in terms of preferred explanations and include the role of those components whose abnormality makes the observation (no longer anomalous) consistent with the description of the system (see [1] and [16]).

This kind of *sentential frameworks* exclusively deals with selective abduction (diagnostic reasoning)[8] and relates to the idea of preserving *consistency*. Exclusively considering the sentential view of abduction does not enable us to say much about creative processes in science, and, therefore, about the nomological and most interesting creative aspects of abduction. It mainly refers to the *selective* (diagnostic) and merely *explanatory* aspects of reasoning and to the idea that abduction is mainly an inference *to the best explanation* (see [15]): when used to express the creativity events it is either empty or replicates the well-known *Gestalt* model of radical innovation. It is empty because the sentential view stops any attempt to analyze the creative processes: the event of creating something new is considered so radical and instantaneous that its irrationality is immediately involved.

For Peirce abduction is an *inferential process* that includes all the operations whereby hypotheses and theories are constructed. Hence abduction has to be considered as a kind of *ampliative* inference that, as already stressed, is not logical and truth preserving: indeed valid deduction does not yield any new information, for example new hypotheses previously unknown.

Usually the sentential models of theoretical abduction are limited, because they do not capture various reasoning tasks [14]:

1. the role of statistical explanations, where what is explained follows only probabilistically and not deductively from the laws and other tools that do the explaining;
2. the sufficient conditions for explanation;
3. the fact that sometimes the explanations consist of the application of schemas that fit a phenomenon into a pattern without realizing a deductive inference;
4. the idea of the existence of high-level kinds of *creative* abductions;
5. the existence of model-based abductions;
6. the fact that explanations usually are not complete but only furnish *partial* accounts of the pertinent evidence (see [34]);
7. the fact that one of the most important virtues of a new scientific hypothesis (or of a scientific theory) is its power of explaining *new*, previously *unknown* facts: "[...] these facts will be [...] unknown at the time of the abduction, and even more so must the auxiliary data which help to explain them be unknown. Hence these future, so far unknown explananda, cannot be among the premises of an abductive inference" (see [9]), observations become real and explainable only by means of new hypotheses and theories, once discovered by abduction.

[8] As previously indicated, it is important to distinguish between *selective* (abduction that merely selects from an encyclopedia of pre-stored hypotheses), and *creative* abduction (abduction that generates new hypotheses).

2.1.2 Model-based abduction

If we want to provide a suitable framework for analyzing the most interesting cases of conceptual changes in science we do not have to limit ourselves to the *sentential* view of theoretical abduction but we have to consider a broader *inferential* one: the *model-based* sides of creative abduction (cf. below).

From the Peirce's philosophical point of view, all thinking is in signs, and signs can be icons, indices or symbols. Moreover, all inference is a form of sign activity, where the word sign includes "feeling, image, conception, and other representation" ([25, 5.283]), and, in Kantian words, all synthetic forms of cognition. That is, a considerable part of the thinking activity is model-based. Of course model-based reasoning acquires its peculiar creative relevance when embedded in abductive processes, so that we can individuate a *model-based abduction*.

Hence, it is in terms of *model-based abduction* (and not in terms of sentential abduction) that we have to think to explain complex processes like scientific conceptual change. Related to the high-level types of scientific conceptual change (see, for instance, [33]) are different varieties of *model-based abductions* (see, for examples, [14]).

Following Nersessian (cf. [22] and [23]), the term "model-based reasoning" is used to indicate the construction and manipulation of various kinds of representations, not mainly sentential and/or formal, but mental and/or related to external mediators.

Although it is controversial as to whether there is any form of representation other than strings of symbols, it is possible, following Johnson-Laird (see [12]), to assume the existence of at least three kinds of *mental* representations:

1. *propositional representations* (strings of symbols such as "the pot is on the table");
2. *mental models* (structural analogs of real world or imagined situations, such as a pot being on a table);
3. *images* (a mental model from a specific perspective, such as looking down on the pot on the table from above).

Obvious examples of model-based reasoning are constructing and manipulating visual representations, thought experiment, analogical reasoning, but also the so-called "tunnel effect" (see [3]), occurring when models are built at the intersection of some operational interpretation domain – with its interpretation capabilities – and a new ill-known domain.

Manipulative abduction [15] - contrasted to theoretical abduction - happens when we are thinking through doing and not only, in a pragmatic sense, about doing. So the idea of manipulative abduction goes beyond the well-known role of experiments as capable of forming new scientific laws by means of the results (the nature's answers to the investigator's question) they present, or of merely playing a predictive role (in confirmation and in falsification). Manipulative abduction refers to an extra-theoretical behavior that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (theoretical) practices. The existence of this kind of extra-theoretical cognitive behavior is also testified by the many everyday situations in which humans are perfectly able to perform very efficacious (and habitual) tasks without the immediate possibility of realizing their conceptual explanation. In the following sections manipulative abduction will be considered from the perspective of the relationship between unexpressed knowledge and external representations.

2.2 Unexpressed knowledge, knowledge creation, and external mediators

The power of model-based abduction mainly depends on its ability to render explicit a certain amount of important information, unexpressed at the level of available data. It also has a fundamental role in the process of transformation of knowledge from its *tacit* to its *explicit* forms, and in the subsequent knowledge elicitation and use. Let us describe how this happens.

As pointed out by M. Polanyi in his epistemological investigation, a large part of knowledge is not explicit, but tacit: we know more than we can tell and we can know nothing without relying upon those things which we may not be able to tell (see [28]). Polanyi's concept of knowledge is based on three main theses: first, discovery cannot be accounted for by a set of articulated rules or algorithms; second, knowledge is public and also to a very great extent personal (i.e. it is constructed by humans and therefore contains emotions, "passions"); third, an important part of knowledge is tacit.

Hence, two levels of knowledge, mutually exclusive but complementary, as they interact in creative tasks, underlie every activity: there is a kind of knowledge we can call *focal*, that is the knowledge about the object or phenomenon in focus; and another kind of knowledge, masked under the first one, and often used as a tool to handle or improve what is in focus, we can call *tacit*. The first one is the knowledge that is transmissible through any systematic language, since it can be relatively easily formulated by means of symbols and it can be digitalized. Tacit knowledge, on the other hand, is characterized by the fact that it is personal, context specific, usually characterized as derived from direct experience, and therefore hard to elicit and communicate. It is a "non-codified, disembodied know-how that is acquired via the informal take-up of learned behavior and procedures" (see [10, p. 92]).

Fleck [5] describes this form of knowledge as "a subtle level of understanding often difficult to put into words, a trained recognition and perception" (p. 119). Tacit knowledge is wholly embodied in the individual, rooted in practice and experience, expressed through skillful execution, and can become useful by means of watching and doing forms of learning and exploitation.

As Polanyi contends, human beings acquire and use knowledge by actively creating and organizing their own experience: tacit knowledge is the practical knowledge used to perform a task. The existence of this kind of not merely theoretical knowing behavior is also testified by the many everyday situations in which humans are perfectly able to perform very efficacious (and habitual) tasks without the immediate possibility of realizing their conceptual explanation. In some cases the conceptual account for doing these things was at one point present in the memory, but now has deteriorated, and it is necessary to reproduce it, in other cases the account has to be constructed for the first time, like in creative experimental settings in science.

Hutchins [11] illustrates the case of a navigation instructor that for 3 years performed an automatized task involving a complicated set of plotting manipulations and procedures. The insight concerning the conceptual relationships between relative and geographic motion came to him suddenly "as lay in his bunk one night". This example explains that many forms of learning can be represented as the result of the capability of giving conceptual and theoretical details to already automatized manipulative executions. The instructor does not discover anything new from the point of view of the objective knowledge about the involved skill, however, we can say that his conceptual awareness is new from the local perspective of his individuality.

We can find a similar situation also in the process of scientific creativity. Too often, in the cognitive view of science, it has been underlined that conceptual change just involves a *theoretical* and "internal" replacement of the main concepts. But usually researchers forget that a large part of this processes are instead due to *practical* and "external" *manipulations* of some kind, prerequisite to the subsequent work of theoretical arrangement and knowledge creation. When these processes are creative we can speak of manipulative abduction (cf. above).

Scientists sometimes need a first "rough" and concrete experience of the world to develop their systems, as a *cognitive-historical* analysis of scientific change (see [21] and [7]) has carefully shown.

The prevailing perspective among philosophers is that the processes of discovery and the consequent new incoming scientific representations are too mysterious to be understood. This view receives support from numerous stories of genius' discoveries, such as Archimedean eureka-experiences. Such accounts neglect periods of intense and often arduous thinking activity, often performed by means of experiments and *manipulative* activity on external objects; these are periods that prepare such "instantaneous" discoveries. It is also important to understand that the scientific process is *complex* and *dynamic*: new representations do not emerge completely codified from the heads of scientists, but are constructed in response to specific problems by the systematic use of heuristic procedures (as pointed out by Herbert Simon's view on the "problem-solving process" [32]).

Traditional examinations of how problem-solving heuristics create new representations in science have analyzed the frequent use of analogical reasoning, imagistic reasoning, and thought experiment, from an internal point of view. However, attention has not been focalized on those particular kinds of heuristics that resort to the existence of *extra-theoretical* ways of thinking (*thinking through doing*, cf. [18]). Indeed many cognitive processes are centered on *external representations*, as a means to create communicable accounts of new experiences ready to be integrated into previously existing systems of experimental and linguistic (theoretical) practices.

It is possible to achieve interesting insights on these problems studying them from a different contrasting approach, which moves away from Simon's paradigm, but which can offer a rational solution to the problem of creativity and conceptual change in terms of mathematical models: the *dynamical* approach [30]. The traditional computational view treats cognition as a process that computes internal symbolic representations of external world. But this approach is considered too reductive, since it is based on the functionalist hypothesis (which cannot render the *external dimension* of cognition), and on a computation of static entities. It is useful to integrate it with a dynamical modeling of cognition, which is able to describe abductive processes as *dynamical entities* "unfolding" in real time (we can also gain a better *cognitive-historical* perspective) (see [20]). From this point of view it is possible to model the terms (objects or propositions) that constitute abduction by considering the *attractors* in a dynamical system. This can be achieved by topologically specifying the *semantic* content of the inferential process through the spatial relations between its defining attractors. We can therefore consider the process of progressive development of "new" concepts and replacement of old ones in terms of temporal evolving patterns defined by interactions between topological configurations of attractors.

Moreover, a central point in the dynamical approach is the importance assigned to the "whole" cognitive system: the cognitive activity is in fact the result of a complex interplay and simultaneous coevolution, in time, of the states of mind, body, and external environment. Even if, of course, a large portion of the complex environment of a thinking agent is internal, and consists of the proper software composed of the knowledge base and of the inferential expertise of the individual, nevertheless a "real" cognitive system is composed by a "distributed cognition" among people and some "external" objects and technical artifacts (see [11] and [24]). For example, in the case of the construction and examination of diagrams in geometrical reasoning, specific experiments serve as states and the implied operators are the manipulations and observations that transform one state into another. The geometrical outcome is dependent upon practices and specific sensory-motor activities performed on a non-symbolic object, which acts as a dedicated external representational medium supporting the various operators at work. There is a kind of an epistemic negotiation between the sensory framework of the geometer and the external reality of the diagram [17]. This process involves an external representation consisting of written symbols and figures that for example are manipulated "by hand". The cognitive system is not merely the mind-brain of the person performing the geometrical task, but the system consisting of the whole body (cognition is *embodied*) of the person plus the external physical representation. In geometrical discovery the whole activity of cognition is located in the system consisting of a human together with diagrams.

An external representation can modify the kind of computation that a human agent uses to reason about a problem: the Roman numeration system eliminates, by means of the external signs, some of the hardest parts of the addition, whereas the Arabic system does the same in the case of the difficult computations in multiplication. The capacity for inner reasoning and thought results from the internalization of the originally external forms of representation. In the case of the external representations we can have various objectified knowledge and structures (like physical symbols – e.g. written symbols, and objects – e.g. three-dimensional models, shapes and dimensions), but also external rules, relations, and constraints incorporated in physical situations (spatial relations of written digits, physical constraints in geometrical diagrams and abacuses) (see [35]). The external representations are contrasted to the internal representations that consist of the knowledge and the structure in memory, as propositions, productions, schemas, neural networks, models, prototypes, images.

The external representations are not merely memory aids: they can give people access to knowledge and skills that are unavailable to internal representations, help researchers to easily identify aspects and to make further inferences, they constrain the range of possible cognitive outcomes in a way that some actions are allowed and other forbidden. The mind is limited because of the restricted range of information processing, the limited power of working memory and attention, the limited speed of some learning and reasoning operations; on the other hand the environment is intricate, because of the huge amount of data, real time requirement, uncertainty factors. Consequently, we have to consider the whole system, consisting of both internal and external representations, and their role in optimizing the whole cognitive performance of the distribution of the various subtasks. It is well-known that in the history of geometry many researchers used internal mental imagery and mental representations of diagrams, but also self-generated diagrams (external) to help their thinking.

3 EXTERNAL REPRESENTATIONS AND MANIPULATIVE ABDUCTION

We have introduced above the notion of *tacit knowledge*. Now we propose an extension of that concept. There is something more important beyond the tacit knowledge "internal" to the subject – considered by Polanyi as personal, embodied and context specific. We can also speak of a sort of tacit information "embodied" into the whole relationship between our mind-body system and suitable external representations. An information we can extract, explicitly develop, and transform in knowledge contents, to solve problems (as it is manifest, for instance, in the geometrical problem contained in the *Meno* [27] (see [15, chapter 1] – even if it philosophers perfectly know that Plato considered this activity just as the result of reminiscence and not of discovery).

As we have already stressed, Peirce considers inferential any cognitive activity whatever, not only conscious abstract thought; he also includes perceptual knowledge and subconscious cognitive activity. For instance in subconscious mental activities visual representations play an immediate role. Peirce gives an interesting example of model-based abduction related to sense activity: "A man can distinguish different textures of cloth by feeling: but not immediately, for he requires to move fingers over the cloth, which shows that he is obliged to compare sensations of one instant with those of another" [25, 5.221]. This surely suggests that abductive movements have also interesting extra-theoretical characters and that there is a role in abductive reasoning for various kinds of manipulations of external objects. *All* knowing is *inferring* and inferring is not instantaneous, it happens in a process that needs an activity of comparisons involving many kinds of models in a more or less considerable lapse of time.

All these considerations suggest, then, that there exist a creative form of thinking through doing,[9] fundamental as much as the theoretical one: *manipulative abduction* (cf. above, and [15, 17]). As already said *manipulative* abduction happens when we are thinking *through* doing and not only, in a pragmatic sense, about doing.

Various *templates* of manipulative behavior exhibit some regularities. The activity of manipulating esternal things and representations is highly conjectural and not immediately explanatory: these templates are "hypotheses of behavior" (creative or already cognitively present in the scientist's mind-body system, and sometimes already applied) that abductively enable a kind of epistemic "doing". Hence, some templates of action and manipulation can be selected in the set of the ones available and pre-stored, others have to be created for the first time to perform the most interesting creative cognitive accomplishments of manipulative abduction.

Some common features of the tacit templates of manipulative abduction, that enable us to manipulate things and experiments in science are related to: 1. sensibility to the aspects of the phenomenon which can be regarded as *curious* or *anomalous*; manipulations have to be able to introduce potential inconsistencies in the received knowledge (Oersted's report of his experiment about electromagnetism is devoted to describe some anomalous aspects that did not depend on any particular theory of the nature of electricity and magnetism); 2. preliminary sensibility to the *dynamical* character of the phenomenon, and not to entities and their properties, common aim of manipulations is to practically reorder the dynamic sequence of events into a static spatial one that should promote a subsequent bird's-eye view (narrative or visual-diagrammatic); 3. referral to experimental manipulations that exploit *artificial apparatus* to free new possible stable and repeatable sources of information about hidden knowledge and constraints (Davy set-up in term of an artifactual tower of needles showed that magnetization was related to orientation and does not require physical contact); 4. various con-

[9] In this way the cognitive task is achieved on *external* representations used in lieu of an internal ones. Here action performs an *epistemic* and not a merely performatory role, relevant to abductive reasoning.

tingent ways of epistemic acting: *looking* from different perspectives, *checking* the different information available, *comparing* subsequent events, *choosing*, *discarding*, *imaging* further manipulations, *re-ordering* and *changing relationships* in the world by implicitly *evaluating* the usefulness of a new order (for instance, to help memory).

Gooding [7] refers to this kind of concrete manipulative reasoning when he illustrates the role in science of the so-called "construals" that embody tacit inferences in procedures that are often apparatus and machine based. The embodiment is of course an expert manipulation of objects in a highly constrained experimental environment, and is directed by abductive movements that imply the strategic application of old and new *templates* of behavior mainly connected with extra-theoretical components, for instance emotional, esthetical, ethical, and economic.

The hypothetical character of construals is clear: they can be developed to examine further chances, or discarded, they are provisional creative organization of experience and some of them become in their turn hypothetical *interpretations* of experience, that is more theory-oriented, their reference is gradually stabilized in terms of established observational practices. Step by step the new interpretation – that at the beginning is completely "practice-laden" – relates to more "theoretical" modes of understanding (narrative, visual, diagrammatic, symbolic, conceptual, simulative), closer to the constructive effects of theoretical abduction. When the reference is stabilized the effects of incommensurability with other stabilized observations can become evident. But it is just the construal of certain phenomena that can be shared by the sustainers of rival theories. Gooding [7] shows how Davy and Faraday could see the same attractive and repulsive actions at work in the phenomena they respectively produced; their discourse and practice as to the role of their construals of phenomena clearly demonstrate they did not inhabit different, incommensurable worlds in some cases. Moreover, the experience is constructed, reconstructed, and distributed across a social network of negotiations among the different scientists by means of construals.

The whole activity of manipulation is in fact devoted to building various external *epistemic mediators*[10] that function as an enormous new source of information and knowledge. Therefore, manipulative abduction represents a kind of redistribution of the epistemic and cognitive effort to manage objects and information that cannot be immediately represented or found internally (for example exploiting the resources of visual imagery).[11]

From the point of view of everyday situations manipulative abductive reasoning and epistemic mediators exhibit very interesting features (we can find the first three in geometrical constructions): 1. action elaborates a *simplification* of the reasoning task and a redistribution of effort across time (see [11]), when we need to manipulate concrete things in order to understand structures which are otherwise too abstract (see [26]), or when we are in presence of *redundant* and unmanageable information; 2. action can be useful in presence of *incomplete* or *inconsistent* information – not only from the "perceptual" point of view – or of a diminished capacity to act upon the world: it is used to get more data to restore coherence and to improve deficient knowledge; 3. action enables us to build *external artifactual models* of task mechanisms instead of the corresponding internal ones, that are adequate to adapt the environment to agent's needs. 4. action as a *control of sense data* illustrates how we can change the position of our body (and/or of the external objects) and how to exploit various kinds of prostheses (Galileo's telescope, technological instruments and interfaces) to get various new kinds of stimulation: action provides some tactile and visual information (e.g., in surgery), otherwise unavailable.

Also natural phenomena can play the role of external artifactual models: under Micronesians' manipulations of their images, the stars acquire a structure that "becomes one of the most important structured representational media of the Micronesian system" [11, p. 172]. The external artifactual models are endowed with functional properties as components of a memory system crossing the boundary between person and environment (for example they are able to transform the tasks involved in allowing simple manipulations that promote further visual inferences at the level of model-based abduction). The cognitive process is *distributed* between a person (or a group of people) and external representation(s), and so obviously *embedded* and *situated* in a society and in a historical culture.

An interesting epistemological situation we have recently studied is the one concerning the role of some special epistemic mediators in the field of non-standard analysis, an "alternative calculus", invented by Abraham Robinson [31], based on infinitesimal numbers in the spirit of Leibniz method. It is a kind of calculus that uses an extension of the real numbers system $\mathbb{R}$ to the system $\mathbb{R}^*$ containing infinitesimals smaller - in the absolute value - than any positive real number. We maintain that in mathematics diagrams play various roles in a typical abductive way. Two of them are central:

- they provide an intuitive and mathematical *explanation* able to help the understanding of concepts difficult to grasp, that appear obscure and/or epistemologically unjustified, or that are *not expressible* from an intuitive point of view;
- they help *create* new previously unknown concepts.

In the construction of mathematical concepts many external representations are exploited, both in terms of diagrams and of symbols. We are interested in our research in diagrams which play an *optical* role – microscopes (that look at the infinitesimally small details), telescopes (that look at infinity), windows (that look at a particular situation), a *mirror* role (to externalize rough mental models), and an *unveiling* role (to help create new and interesting mathematical concepts, theories, and structures).[12]

Optical diagrams play a fundamental explanatory (and didactic) role in removing obstacles and obscurities and in enhancing mathematical knowledge of critical situations. They facilitate new internal representations and new symbolic-propositional achievements. In the example we have studied in the area of the calculus, the extraordinary role of the optical diagrams in the interplay standard/non-standard analysis is emphasized. In the case of our non-standard analysis examples, some new diagrams (microscopes within microscopes) provide new mental representations of the concept of tangent line at the infinitesimally small regions. Hence, external representations which

[10] This expression, introduced by Magnani [15], is derived from the cognitive anthropologist Hutchins (see [11]), who coined the expression "mediating structure" to refer to various external tools that can be built to cognitively help the activity of navigating in modern but also in "primitive" settings. Any written procedure is a simple example of a cognitive "mediating structure" with possible cognitive aims, so mathematical symbols and diagrams: "Language, cultural knowledge, mental models, arithmetic procedures, and rules of logic are all mediating structures too. So are traffic lights, supermarkets layouts, and the contexts we arrange for one another's behavior. Mediating structures can be embodied in artifacts, in ideas, in systems of social interactions [...]" [11, pp. 290–291].

[11] It is difficult to preserve precise spatial and geometrical relationships using mental imagery, in many situations, especially when one set of them has to be moved relative to another.

[12] The epistemic and cognitive role of mirror and unveiling diagrams in the discovery of non-Euclidean geometry is illustrated in [17].

play an "optical" role can be used to allow us a better understanding of many critical mathematical situations and, in some cases, to discover (or rediscover) more easily and nicely sophisticated properties. The role of an "optical microscope" that shows the behavior of a tangent line is illuminating. In standard analysis, the change dy in y along the tangent line is only an approximation of the change Δy in y along the curve. But through an optical microscope, that shows infinitesimal details, we can see that $dy = \Delta y$, and then the quotient $\Delta y/\Delta x$ is equal to the quotient dy/dx when $dx = \Delta x$ is infinitesimal (see Figure 1 and, for details, [19]). This eliminates some difficulties of the representation of the tangent line as limit of secants, and introduces a more intuitive conceptualization: the tangent line "merges" with the curve in an infinitesimal neighborhood of the contact point.

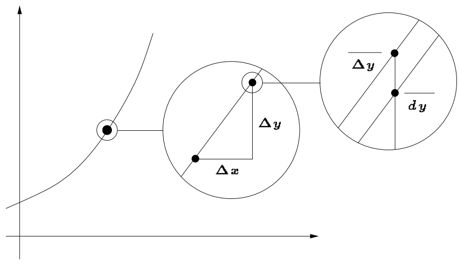

Figure 1. An optical diagram shows an infinitesimal neighborhood of the graph of a real function.

Some diagrams could also play an unveiling role, providing new light on mathematical structures: it can be hypothesized that these diagrams can lead to further interesting creative results. The optical and unveiling diagrammatic representation of mathematical structures activates *direct perceptual operations* (for example identifying how a real function appears in its points and/or to infinity; how to really reach its limits).

We stated that in mathematics diagrams play various roles in a typical abductive way. We can add that:

- they are *epistemic mediators* able to perform various abductive tasks in so far as
- they are *external representations* which are devoted to provide explanatory abductive results.

4 CONCLUSION

It is clear that the manipulation of external objects helps human beings in their creative tasks. We have illustrated the strategic role played by the so-called traditional concept of "implicit knowledge" in terms of the recent cognitive and epistemological concept of manipulative abduction, considered as a particular kind of abduction that exploits external models endowed with delegated (sometimes implicit) cognitive roles and attributes. Abductive manipulations operate on models that are external and the strategy that organizes the manipulations is unknown *a priori*. In the case of creative manipulations of course the result achieved is also *new*, and adds properties not contained before.

If we simply look at the birth of modern science, an epistemological situation corresponding to that one delineated in the previous section is immediately evident. Already in the *Dialogues Concerning the Two Chief World Systems* (see [6]), accentuating the role of observational manipulations, Galileo presents an anatomist that, "manipulating" a cadaver, is able to get new, not speculative, information that goes beyond the "world of paper" of the Peripatetic philosophy. A lot of new information in science is reached by observations and experiments, and experiments are the fruit of various kinds of artifactual manipulations: the different strategies correspond to the expert manipulations of objects in a highly constrained experimental environment, directed by *abductive* movements that imply the application of old and new extra-theoretical templates of behavior. Peirce too, again, underlies this point speaking about mathematical and geometrical reasoning that "consists in constructing a diagram according to a general precept, in observing certain relations between parts of that diagram not explicitly required by the precept, showing that these relations will hold for all such diagrams, and in formulating this conclusion in general terms" [25, 1.54].

In scientific practice there is a lot of procedural, extra-sentential and extra-theoretical aspects indispensable to provide knowledge and information otherwise hard to grasp. By making explicit them we can rationally and positively integrate the previously existing scientific encyclopedia. The enhancement of the analysis of these important human skills can increase knowledge on inferences involving creative, analogical, spatial, and simulative aspects, both in science and everyday situations, so that this can extend the epistemological, computational, and the psychological theory.

REFERENCES

[1] C. Boutilier and V. Becher, 'Abduction as belief revision', *Artificial Intelligence*, **77**, 43–94, (1995).

[2] T. Bylander, D. Allemang, M.C. Tanner, and J.R. Josephson, 'The computational complexity of abduction', *Artificial Intelligence*, **49**, 25–60, (1991).

[3] A. Cornuéjols, A. Tiberghien, and G. Collet, 'A new mechanism for transfer between conceptual domains in scientific discovery and education', *Foundations of Science*, **5**(2), 129–155, (2000). Special Issue on 'Model-based Reasoning in Science: Learning and Discovery', ed. by L. Magnani, N.J. Nersessian, and P. Thagard.

[4] P. Flach and A. Kakas, eds. *Abductive and Inductive Reasoning: Essays on Their Relation and Integration*, Dordrecht, 2000. Kluwer Academic Publishers.

[5] J. Fleck, 'Informal information flow and the nature of expertise in financial service', *International Journal of Technology Management*, **11**(1–2), 104–128, (1996).

[6] G. Galilei, 'Dialogues concerning the two chief world systems (1632)', in *The Scientific Background to Modern Philosophy*, ed., M.R. Matthews, Hackett, Indianapolis/Cambridge, (1989). Translated by S. Drake.

[7] D. Gooding, *Experiment and the Making of Meaning*, Kluwer, Dordrecht, 1990.

[8] C.G. Hempel, *Philosophy of Natural Science*, Prentice-Hall, Englewood Cliffs, NJ, 1966.

[9] J. Hintikka, 'What is abduction? The fundamental problem of contemporary epistemology', *Transactions of the Charles S. Peirce Society*, **34**, 503–533, (1998).

[10] J. Howells, 'Tacit knowledge, innovation and technology transfer', *Technology Analysis & Strategic Management*, **8**(2), 91–105, (1996).

[11] E. Hutchins, *Cognition in the Wild*, MIT Press, Cambridge, MA, 1995.

[12] P.N. Johnson-Laird, *Mental Models*, Harvard University Press, Cambridge, MA, 1983.

[13] T.S. Kuhn, *The Structures of Scientific Revolutions*, University of Chicago Press, Chicago, 1962.

[14] L. Magnani, 'Inconsistencies and creative abduction in science', in *AI and Scientific Creativity. Proceedings of the AISB99 Symposium on Scientific Creativity*, pp. 1–8, Edinburgh, (1999). Society for the Study of Artificial Intelligence and Simulation of Behaviour, University of Edinburgh.

[15] L. Magnani, *Abduction, Reason, and Science. Processes of Discovery and Explanation*, Kluwer Academic/Plenum Publishers, New York, 2001.

[16] L. Magnani, 'Limitations of recent formal models of abductive reasoning', in *Abductive Reasoning. KKR-2, 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)*, eds., A. Kakas and F. Toni, pp. 34–40, Berlin, (2001). Springer.

[17] L. Magnani, 'Epistemic mediators and model-based discovery in science', in *Model-Based Reasoning: Science, Technology, Values*, eds., L. Magnani and N.J. Nersessian, pp. 305–329, New York, (2002). Kluwer Academic/Plenum Publishers.

[18] L. Magnani, 'Thinking through doing, external representations in abductive reasoning', in *AISB 2002 Symposium on AI and Creativity in Arts and Science*, London, (2002). Imperial College.

[19] L. Magnani and R. Dossena, 'Perceiving the infinite and the infinitesimal world: unveiling and optical diagrams and the construction of mathematical concepts', Technical Report, University of Pavia, Italy, (2002).

[20] L. Magnani and M. Piazza. Morphodynamical abduction: causation by attractors dynamics of explanatory hypotheses in science, 2002. International Conference *Causation and Explanation in Natural and Social Science (CENSS)*.

[21] N.J. Nersessian, 'How do scientists think? Capturing the dynamics of conceptual change in science', in *Cognitive Models of Science*, ed., R. Giere, Minnesota Studies in the Philosophy of Science, pp. 3–44, Minneapolis, (1992). University of Minnesota Press.

[22] N.J. Nersessian, 'Should physicists preach what they practice? Constructive modeling in doing and learning physics', *Science and Education*, **4**, 203–226, (1995).

[23] N.J. Nersessian, 'Model-based reasoning in conceptual change', in *Model-based Reasoning in Scientific Discovery*, eds., N.J. Nersessian, L. Magnani, and P. Thagard, pp. 5–22, New York, (1999). Kluwer Academic/Plenum Publishers.

[24] D.A. Norman, *Things that Make Us Smart. Defending Human Attributes in the Age of the Machine*, Addison-Wesley, Reading, MA, 1993.

[25] C.S. Peirce, *Collected Papers*, Harvard University Press, Cambridge, MA, 1931–58. 1–6, ed. by C. Hartshorne and P. Weiss, 7-8, ed. by A.W. Burks.

[26] J. Piaget, *Adaption and Intelligence*, University of Chicago Press, Chicago, 1974.

[27] Plato, *Plato in Twelve Volumes*, Harvard University Press, Cambridge, MA, 1977. vol. II, Laches, Protagoras, Meno, Euthydemus, with an English translation by W.R.M. Lamb.

[28] M. Polanyi, *The Tacit Dimension*, Routledge & Kegan Paul, London, 1966.

[29] K.R. Popper, *The Logic of Scientific Discovery*, Hutchinson, London, New York, 1959.

[30] R.F. Port and T. van Gelder, eds. *Mind as Motion. Explorations in the Dynamics of Cognition*, Cambridge, MA, 1995. MIT Press.

[31] A. Robinson, *Non-Standard Analysis*, North Holland, Amsterdam, 1966.

[32] H.A. Simon, *Models of Discovery and Other Topics in the Methods of Science*, Reidel, Dordrecht, 1977.

[33] P. Thagard, *Conceptual Revolutions*, Princeton University Press, Princeton, 1992.

[34] P. Thagard and C.P. Shelley, 'Abductive reasoning: logic, visual thinking, and coherence', in *Logic and Scientific Methods*, eds., M.L. Dalla Chiara, K. Doets, D. Mundici, and J. van Benthem, pp. 413–427, Dordrecht, (1997). Kluwer.

[35] J. Zhang, 'The nature of external representations in problem-solving', *Cognitive Science*, **21**(2), 179–217, (1997).