Morphodynamical Abduction.
Causation by Attractors Dynamics of Explanatory Hypotheses in Science

Lorenzo Magnani and Matteo Piazza
Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy, and Georgia Institute of Technology, Atlanta, USA,
lmagnani@cc.gatech.edu,mpiazza@pure-reason.unipv.it

Abstract: Philosophers of science today by and large reject the cataclysmic and irrational interpretation of the scientific enterprise claimed by Kuhn. Many computational models have been implemented to rationally study the conceptual change in science. In this recent tradition a key role is played by the concept of *abduction* as a mechanism by which new explanatory hypotheses are introduced. Some problems in describing the most interesting abductive issues rise from the classical computational approach. It describes a cognitive process (and so abduction) by the manipulation of internal symbolic representations of external world. This view assumes a discrete set of representations fixed in discrete time jumps, and cannot account for the issue of *anticipation* and *causation* of a new hypothesis. An integration of the traditional computational view with some ideas developed inside the so-called *dynamical* approach can suggest some important insights. The concept of *attractor* is very significant. It permits a description of the abductive generation of new hypotheses in terms of a catastrophic rearrangement of the parameters responsible for the behavior of the system.

1. INTRODUCTION

A full theory of human cognition requires understanding the nature of scientific reasoning: one of the most abstract and creative kinds of inference. In particular we need an understanding of those processes underlying creative reasoning, and conceptual change in science.

Many philosophical efforts in the last century have been spent to study the conceptual change in science. In the mid–1960s many critics challenged the comforting picture of conceptual change in terms of continuous and cumulative steps. Contrary to this picture Kuhn claimed that conceptual change in science is analyzable as a kind of irrational and obscure *Gestalt-switch*, that accounts for the inventive processes and the achievement of new scientific theories and paradigms (Kuhn, 1962). Kuhn argued that major changes in science are best characterized as *revolutions*, involving overthrow and replacement of the reigning conceptual systems and world views by means of new ones *incommensurable* with them. Kuhn brought philosophers of science to distinguish between the logic of discovery and the logic of justification (i.e. the distinction between the psychological side of creation and the logic argument of proving new discovered ideas by facts). The consequent conclusion was that a logic of discovery (and a *rational* model of discovery) does not exist: scientific change is cataclysmic and irrational, dramatic and discontinuous.

Today philosophers of science have abandoned this attitude. The researchers who work on scientific change tend now to stress attention on the problem of rational choice between competing theories and hypotheses and the discovery processes. This also **leads** to the problem of understanding how scientists combine their individual human cognitive abilities with the conceptual resources available to them as members of a scientific community and of a wider social context. It is by means of this synthesis that the creation, elaboration, and communication of a new emerging representation of a scientific domain is made possible.

Many researchers in the area of cognitive sciences consider scientific thinking, (and thinking activity in general), as related to a kind of "representational" system that we can implement in a computational model: thinking is a form of computation. Following the idea that a full understanding of mental processes is possible only by a computational implementation (Johnson-Laird, 1983), these models have been implemented in AI programs where data structures and procedures correspond to assumed mental structures and processes (Thagard, 1992).

To better understand the complex problem of conceptual change in science, besides the ideas elaborated in the AI areas of knowledge representation, problem solving, and machine learning, we need the important concept of *abduction*. Scientific theories contain many theoretical hypotheses that cannot be built by simple generalization of observations. Indeed, Peirce presented abduction as a mechanism by which it is possible to account for the generation of new explanatory hypotheses in science.

2. THE ABDUCTIVE FRAMEWORK

What is abduction? If we see a broken horizontal glass on the floor (an *anomaly* that needs to be solved/explained), we might explain this fact by postulating the effect of wind blowing shortly before: this is not certainly a deductive consequence of the glass being broken (a cat may well have been responsible for it). 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 (Magnani, 2001): 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[1] 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*[2] 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.

The type of inference called abduction was studied by Aristotelian syllogistics, as a form of α1παγωγη¿, and later on by mediaeval reworkers of

[1] Magnani (2001) 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.

[2] Manipulative abduction and epistemic mediators are introduced and illustrated in Magnani (2001).

syllogism. A hundred years ago, Peirce interpreted abduction essentially as an “inferential” *creative process* of generating a new hypothesis, as the only means to a real objective knowledge improvement, besides the well-known deduction and induction.

Since the time of John Stuart Mill (1843), 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 (1955) 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.

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[3].

(This syllogism is known as Barbara).

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

[3] The famous syllogistic example given by Peirce is:

1. All beans from this bag are white.
2. These beans are from this bag.
3. These beans are white.

(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)[4]. Starting from knowing – selecting – (1) and "observing" (3) we can infer (2) by performing a selective abduction[5]. The abductive inference rule corresponds to the well-known fallacy called affirming the consequent (simplified to the propositional case)

$$\frac{\varphi \rightarrow \psi \quad \psi}{\varphi}$$

Thus, *selective abduction* is the making of a preliminary guess that introduces a set of plausible diagnostic hypotheses, followed by deduction to explore their consequences, and by induction to test them with available patient data, (1) to increase the likelihood of a hypothesis by noting evidence explained by that one, rather than by competing hypotheses, or (2) to refute all but one.

If during this first cycle new information emerges, hypotheses not previously considered can be suggested and a new cycle takes place. In this case the *nonmonotonic* character of abductive reasoning is clear and arises from the logical unsoundness of the inference rule: it draws defeasible conclusions from incomplete information. All recent logical accounts ("deductive") concerning abduction have pointed out that it is a form of nonmonotonic reasoning. It is important to allow the guessing of explanations for a situation, in order to discount and abandon old hypotheses, so as to enable the tentative adoption of new ones, when new information about the situation makes them no longer the best.

From the epistemological point of view abductive reasoning can be employed to explain *anomalous* data, creating new explanatory hypotheses according to the general schema:

> Anomalous phenomenon $G(a)$ has to be explained (that is, why a is G?).
> The general rule $(x)(Fx \rightarrow Gx)$, that is any F is G, could explain $G(a)$.
> Then, may be $F(a)$, that is a could be F.

[4] We can consider this inference a sort of generalization from a sample of patients [or of beans] to the whole population of them [or of beans in the bag].

[5] We have to remark that at the level of the syllogistic treatment of the subject Peirce calls this kind of argumentation "hypothesis"; he will introduce the term abduction only in his later theory.

Hence, abduction is that process that enables inference of certain facts or hypotheses able to explain anomalous or unknown phenomena. Peirce stated that abduction has a purely creative value. As a hypothesis, it cannot be justified: when a scientific hypothesis is formulated for the first time and then empirically corroborated, it possesses the status of real discovery, an innovation that establishes a new worldview. It is the case, for example, of Kepler's hypothesis of the elliptic planetary orbit.

Many attempts have been made to model abduction by developing some formal tools in order to illustrate its relationships with the different forms of deductive reasoning (Bylander et al., 1991). Some of these models are based on the theory of the *epistemic state* of an agent (Boutilier and Becher, 1995), 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*).

This kind of *sentential* framework exclusively deals with selective abduction (diagnostic reasoning)[6] 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. It mainly refers to the *selective* (diagnostic) and *explanatory* aspects of reasoning and to the idea that abduction is mainly an inference to the best explanation: 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. But, as stated by Peirce, abduction is an *inferential process* that includes all the operations whereby hypotheses and theories are constructed. Abduction has, then, to be considered as a kind of *ampliative* inference that is not logical and truth preserving (in the sense of deductive). It provides, sometimes, a very radical new perspective: indeed valid deduction does not yield any new information, for example previously unknown new hypotheses.

2.1 Model-based and manipulative 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.

From 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

[6] We have to distinguish between selective and creative abduction. Abduction that merely *selects* from an encyclopedia of pre-stored hypotheses is called selective. Abduction that generates *new* hypotheses is called creative (see Magnani, 2001).

other representation" (*CP*, 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*. For example, it is well known the importance Peirce ascribed to diagrammatic thinking, as shown by his discovery of the system of predicate logic based on "existential graph". The interesting thing is, however, that Peirce considers to be inferential any cognitive activity whatever, not only conscious abstract thought; he also includes perceptual knowledge and subconscious cognitive activity. Perception is viewed by Peirce as a fast and uncontrolled knowledge-production procedure. Perception, in fact, is a vehicle for the instantaneous retrieval of knowledge that was previously structured in our mind through inferential processes. Peirce gives a meaningful example of model-based abduction related to sense activity: "A man can distinguish different textures of cloth by feeling: but not immediately, for he must move fingers over the cloth, which shows that he is obliged to compare sensations of one instant with those of another" (*CP*, 5.221). This surely suggests two things: 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.

It is in terms of *model-based abduction* (and not in terms of sentential abduction) that we have to think of explaining complex processes like scientific conceptual change. There are different varieties of *model-based abductions* (see, for example, Magnani 1999) related to the high-level types of scientific conceptual change (see, for instance, Thagard, 1992).

Following Nersessian (cf.. Nersessian 1995 and 1999), 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. Obvious examples of model-based reasoning are constructing and manipulating visual representations, thought experiment, and analogical reasoning. The centrality of *mental* modeling to cognition is testified by many accounts (see, e.g., Magnani, Nersessian and Thagard, 1999, Magnani and Nersessian, 2002).

Although there is still much to learn about the cognitive processes underlying modeling, this hypothesis is attractive, because it provides a cognitive foundation for taking seriously the productive modeling practices of scientists as the reasoning through which new conceptual structures are constructed. "Modeling is not at all ancillary to doing science, but central to constructing accounts of the natural world" (Giere, 1999).

Manipulative abduction (Magnani, 2001) – in contrast 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 (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.

3. ABDUCTION AS EMBODIED COGNITION

As we have seen, for Peirce extra-theoretical aspects and manipulations of "external" objects in reasoning are also important. Paying attention to the *perceptual* and *manipulative* dimension of cognition, Peirce reminds us that real cognitive systems are not "isolated" and autonomous entities. Cognition is *embodied*, and the interactions between mind and external environment are its central aspects. Knowledge is possible only by means of a constant exchange of information in a complex *distributed* system that crosses the boundary between the person and the surrounding environment.

This aspect has not been sufficiently stressed. Traditional accounts in cognitive science in fact describe reasoning processes in terms of the so-called "computational" approach. The main idea is that cognition is the operation of a special mental "computer", located in the brain. Sensory organs deliver up to the computer representations of the environment that have to be elaborated. Representations are *static* structures of *discrete* symbols; cognitive operations are transformations from one static symbol structure to the next; these transformations are discrete, effectively *instantaneous*, and sequential (see Port and Van Gelder, 1995). This approach, because of the *functionalist* hypothesis, cannot adequately render the *embodied* dimension of cognition.

Another central problem is *time*. Cognitive processes, as complex structures in natural systems, unfold continuously and simultaneously in real time, while computational models specify a discrete sequence of states in arbitrary time periods. The notion of time is present in any intelligent activity (and so it is fundamental in abductive reasoning): time is profoundly involved in human perception and understanding of the world. Things appear to remain in a particular state for a period of time until a certain event happens, so we can say that time is central to reasoning about change and action (as we have seen in the case of abduction related to sense activity). The first task is to consider the different states or conditions of a thing and define how they are related; we can say that we have knowledge about a *causal* relationship when we can use it to define how the related system evolves. Temporal reference is highly integrated in scientific knowledge but this is also the case in human common sense reasoning, both verbal and at the level of visual imagery. Humans are involved in managing time when coordinating with the environment; memories and many mental models seem to be organized around time - past events come to mind when reconstructed with the help of a time framework (chronologically). An alternative approach, then, should be founded on the main argument that cognitive processes happen in time, and are constantly related to the external world.

3.1 Dynamical systems

There is a theory able to describe the behavior of natural systems in time: the *dynamical systems theory*, a widely used, powerful, and successful descriptive framework in natural science. We can use the mathematical tools of dynamics to study cognition by thinking to a cognitive system not just as a computer, but as a dynamical system, made of mind, body, and external environment, mutually and simultaneously influencing and coevolving. To clarify some hidden processes in creative (abductive) reasoning we can try to integrate this dynamical perspective and the computational one.

The attribute "dynamical" refers to the way a system evolves (or behaves) in time. A "system" is a collection of correlated parts perceived as a single entity. Then, a dynamical system is a system changing in time. It is the *state* of the system that changes: that is, the overall look of the system in a certain instant. A dynamical system is defined by the *space* of all possible states it can assume (*state space*), and by a rule (the *dynamic*), which determines the system evolution in a given instant with respect to any initial state (Hirsch, 1984). It is possible to study the system behavior by analyzing change in its states.

In history of science, the most important dynamical system is the solar system. Sun, stars, and moon are the parts making up the system; configurations and speeds they can assume make up the states. The main problem is to find the right dynamics by which it is possible to describe the evolution of the system, and to make predictions (for example in the case of eclipses). Early descriptions (Tolomeus, Copernicus, Brahe, Kepler) used mathematical models to study astronomical dynamics; then, beginning with Galileo, Newton and Leibniz, *differential equations* became the main mathematical means used in describing systems under pressure of forces responsible of its evolution. These equations (in which both function of variables and their derivatives appear) are able to specify how a system evolves, in any moment, in function of a given state (with an accentuated *sensibility to initial conditions*). For instance, the differential equation

$$\ddot{x} = -\frac{k}{m}x \qquad (m\ddot{x} + kx = 0)$$

describes the way an object with mass, hung from a spring, rebounds, defining the instantaneous acceleration ($\ddot{x}$) in function of its position (x); k and m are the constants relating to the tension of the spring and to the mass.

If a system can be described dynamically, it means it has n characteristics (e.g. position, mass, etc.) evolving simultaneously in time. These characteristics can be measured, in any given instant, and associated to a real number. The overall state of the system can then be thought of as an ordered set made of n real numbers, and the state space as isomorphic to a space of real numbers, the n dimensions of which correspond to the different system characteristics (the *phase space*). The evolution of the system in time corresponds to a sequence of points, a *trajectory*, inside the phase space. This sequence can usually be described mathematically as a function of time, considered an independent variable, giving a solution to the system of differential equations. A definition of "dynamical system" can then be a state-determined system with a numerical phase space and an evolution rule which can define trajectories in the space.

3.2 Attractors

Some dynamical systems are so complex, behaving non-linearly and erratically, jumping from a point in the state space to another very different in a brief time (like in the case of the states of the atmosphere). However, notwithstanding these sudden changes, a dynamical system has a series of states, the so-called *attractors*, which tend to remain stable (Figure 1). A system can have a lot of attractors, contemplating more than a single stable state. The transition from one attractor to another is called a *phase transition* (as in the case of water that is so cold that it become ice: in these cases, small local changes, actually lead the system to a qualitatively different state).

Figure 1. An intuitive representation of the idea underlying the concept of attractor is useful in clarifying the idea. Think of a marble rolling on a plane as far as it falls in a hollow like that in the picture. The marble will rotate inside it; then it will reach the resting position, at the bottom. Attractor is the stationary point corresponding to that position.

In the last section we will suggest a description of abduction in terms of attractors: creative abduction can be seen as the formation of a new attractor, so that attractors will represent the cognitive system tendencies to produce interpretive models.

As for now it is interesting to understand how attractors evolve and change as the system parameters gradually change. Dynamical systems can constitute the background in the dynamical approach to cognitive science: the difference between dynamical and computational systems is notable. Computational systems have states made by configurations of symbols and evolution rules, which specify the transformations between two different symbolic configurations (Figure 2).

Numerical phase spaces possess a metric by which it is possible to determine the distance among points. If the phase space is dense enough, then it is possible to find a set of other points between two points, describing the state of the system in any given instant (the notion of time used to describe the system has the same properties of the *real* time).

This kind of description is not possible from a classical computational point of view: in this last case there is not a natural notion which can define the distance between two states of the system, and the concept of “time” is just a synonym of “order” (t_1, t_2, …). Then, it is not possible to specify the direction and the speed in the system evolution. But it is important to determine it, because real processes (including cognitive processes), in real world, occur in real time at a certain speed. Time is essential.

Figure 2. Computational and dynamical modeling of cognition.

3.3 Cognitive processes as super-representational

Usually cognitive scientists distinguish cognitive processes from other natural processes by assuming they are dependent on *knowledge*, something that can be stored and used. Computational approach is centered on the idea that knowledge is in some way "represented", and that cognitive processes are nothing but manipulations of representations.

As already illustrated above, mental representations are usually conceived as symbolic, in the sense that they have a combinatorial syntax and compositional semantics; that is, mental processes are "structure-sensible" because they are defined by the implemented symbolic structure (Fodor and Pylyshyn, 1988). Hence, it seems, there is a problem: how is it possible for a dynamical model, which does not require the notion of "representation", to describe cognitive processes, since they depend on knowledge? The fact that these models are not founded on manipulation of symbolic structures does not means they do not admit them at all. Cognitive processes can be considered *super-representational*: they not only involve some kind of symbolic representations, but much more. They are not so simple that we can describe them without the notion of representation: they are so complex that the simple notion of representation only constitutes a part in the global description of their functioning. Anyway, a representational state it associable to many aspects of dynamical models: states, attractors, trajectories, bifurcations (sudden qualitative transformations). In a computational model, the rules governing the system are defined on the basis of the entities with a representational status, whereas in a dynamical one rules are defined by numerical states. This means that a dynamical system can be "representational" even if its evolution rules are not specified by representations (for example, "representations" of objects stored in memory are nothing but configurations of attractors in the phase space of the system). Then, also a dynamical system is able to store knowledge which influences its behavior.

3.4 Embodied cognition and qualitative modeling

We have already seen that knowledge also involves pragmatic and "embodied" aspects. Advocates of computational approach claim that the cognitive system is made by the mind, considered a sort of control unit embedded in a body in turn embedded in an external environment. So the cognitive system interacts with the external world through the body: there are transducers which translate the physical interaction between body and environment in states defined by symbolic representations, medium of cognitive processes. What it is important is that since the system works through representations, it is possible to consider it as autonomous, isolated with respect to the body and the external world, the function of which is simply to transform representations (inputs) into other representations (outputs).[7]

On the other hand, dynamical systems are defined by multiple aspects, which simultaneously evolve in a temporal continuum and affect each other. Since nervous system, body and environment are continuously evolving, and affect each other, the cognitive system cannot be constituted simply by an isolated "brain": it is a single system in which the three subsystems are unified. The system does not interact cyclically with body and environment by means of symbolical inputs and outputs: internal and external processes are *coupled* in a way so that they interact constantly. Moreover, the process is not sequential, because all the aspects of the system constantly change simultaneously.

Surely there are many cognitive performances that exhibit a sequential behavior: for example the pronunciation of a sentence. But this "sequential aspect" is nothing but something that emerges in time, underlying the overall trajectory of the system, in which the rules of evolution do not specify a sequential change, but a mutual and simultaneous co-evolution. Hence it is natural to think of the system evolution in terms of its movement inside the state space. Consequently the phase space is described numerically, it is possible to apply the notion of *distance*: an interesting point in the dynamical theory is just the fact that it permits to *geometrically* conceptualize cognitive processes. The distinctive character of a cognitive process developing in time depends on the disposition in space of the states through which it passes.

[7] The idea is directly derived from the notion of "functionalism".

Notwithstanding the difficult in fulfilling the main purpose of the dynamical approach (that is to build a *quantitative* model of cognitive processes), it is nevertheless interesting to develop mathematical models which express *qualitatively* similar (and not quantitatively exact) behaviors to the studied phenomena. Qualitative modeling aims at building qualitative causal descriptions of the studied systems. A qualitative model represents the *structure* of a system, and gives a qualitative description of its overall behavior. Some dynamical properties, such as catastrophic jumps, oscillations, and chaotic behaviors due to variations in control parameters, can be examined without knowing the exact equation governing the system evolution, also thanks to the "geometrical" properties exhibited by this kind of modeling. The description of cognitive systems through attractors accords with the "compositional" aspects of qualitative modeling: the structure is represented as a set of devices interconnected by causal principles; the study of the relation between these devices permits analysis of the state of the system.

4. MORPHODYNAMICAL ABDUCTION AND ADUMBRATIONS

Morphodynamical abduction is abduction considered in the light of the geometrical framework described above. The main idea is that a complex system, as the cognitive one, and its transformations, can be described in terms of a configurational structure. That is, different mental states are defined by their geometrical relationships within a larger dynamical environment. This suggests that the system, in any given instant, possesses a general *morphology* we can study by observing how it changes and develops. The term *morphodynamics* refers to those theories aimed at explaining morphologies and iconic, schematic, *Gestalt*-like aspects of structures, whatever their underlying physical substrate may be, by using the mathematical theory of dynamical systems (see Thom, 1980).

As we have already said, our aim is to cognitively understand what happens during the process of "creation" of an interpretive model, when the mind finds an order in the disordered flow of experience. We maintain that it is possible to obtain interesting suggestions by integrating the traditional computational models with some ideas coming from the dynamical approach entangled in the tradition of phenomenology (cf. below).[8]

As we have illustrated above by introducing the notion of trajectory, it is possible to represent the evolution of a system in a diagram by "drawing" its

[8] Working then in the paradigm of the so-called *naturalized phenomenology*, an approach that is aimed at supporting phenomenology with scientific explanations (neurophysiological, mathematical, physical, etc.) (see Petitot and others, 1999).

function in a n-dimensional space (in which n represents the number of the important parameters influencing the system), as in Figure 3. In the picture it is possible to see a simple system evolving in time (considered as an independent variable).

Figure 3. Representation of the evolution of the system in time

This diagram is useful to understand the cognitive processes underlying abduction from a dynamical point of view. We can then extend the idea to the entire mechanism of conceptual change.

The system remains in a stable state, the initial state S_i, until the parameters by which it is influenced lead it to some unstable state (in the picture, at t_1 the parameter x assumes a certain critical value). It is a catastrophic rearrangement inside the overall aspect of the system which changes the initial state into the final one S_f. The system rests inside two attractors between the points *a,b* and *c,d*, while it goes through unstable configurations that immediately disappear between *b* and *c*.

This model underlies the fact that the parameters involved (and their interactions) determine the behavior of a cognitive system. We can in fact represent the system as in Figure 4. By using a metaphor, we can consider the parameters (P_1, P_2 , …) as "interacting" among themselves as "atoms under forces" (F_1, F_2, …). Any parameter acts on the other ones, moving the overall cognitive "structure". Suddenly, in a certain instant, this activity stops, and the system reaches stability. The values the parameters assume determine the overall configuration of the system in a given instant.

It is easy to understand in these terms the evolution represented in Figure 3. The parameters maintain the system in the same qualitative state until in *b*, at t_1, some values disturb the overall equilibrium. That leads to a new qualitative state in the system.

Figure 4. Interactions between parameters.

We have used an intuitive representation of the concept of attractor in the previous paragraph. This was justified by identifying possible mental states with attractors in the state space (considered as a geometrical surface in which possible mental states interact) of the cognitive system. Like in the case of the intuitive representation of the relativistic conception of gravitation, we can see this surface as a flat horizontal rubber sheet (Figure 6a). The attractor corresponds to the zone in which we can imagine placing a large and heavy sphere. Its weight will stretch the sheet down and distort the system (Figure 6b and 6c). Therefore, if we imagine the behavior of the cognitive system as a small ball that moves inside the rubber sheet, touching different points over the overall surface, we can easily see how the structure, the "shape" of the space, affects its motion. The parameters responsible for the behavior of the system determine the "weight" of the attractor, then the shape of the surface. We maintain that this process is assimilable to the notion of *anticipation* (see below) developed in Husserl's phenomenology.

By using a particular form of diagram playing an *optical* role (see Magnani and Dossena, 2002) – able to look and *unveil* any given instant in the evolution of the system – it is possible to understand how the system behaves in purely transitory instants, or when it reaches a stable position because of an attractor. Here it is interesting to combine Figure 1 with some optical diagrams working as "microscopes" (Figure 5). This enables us to combine a "dynamical" representation of the system with some "snapshots" of its single states in given instants. Figure 1 can be seen, temporarily, as describing the movements inside the states-surface system. If we represent the important parameters as in Figure 4, we can identify any point of the picture in Figure 1 with their different settings in time.

Figure 5. A diagram working as a "microscope". It shows how the interaction between the parameters of the system leads to a stable state (the attractor).

The system can assume any unstable configuration possible in that space, until it reaches the bottom (that is, configuration *C*). We can analyze each one of those states by using the "microscope" on any point in the space. We see in Figure 5, for instance, two points, *A* and *B*, corresponding to two different unstable states, and the final stable point *C*. In terms of the metaphor used above, we claim that the parameters interact between themselves "folding up" the surface, until they reach a position of equilibrium: the heaviest one, that gives birth to the attractive zone. In this way, in order to understand the cognitive behavior of the system, it is possible to study the overall behavior of the system looking at any interesting space "shape" that leads it from point *a* to point *c*, passing through point *b* (see Figure 6).

Figure 6. Deformations in the cognitive space of the system.

From state *A* to *B* the space begins to be subject to a first deformation. *A* represents a *tabula rasa* state, when the agent is still searching for, and no ideas influence its judgment. Suddenly something changes in the parameters, leading to the transitory state *B*. This state is just transitory and unstable: at this point a *bifurcation* happens. In fact, the system can "reject" state *B* by coming back to state *A* (the weight of parameters is not yet irremediably heavy), or any other previous state. Otherwise, the system can persist in state *B*, anticipating the abductive inference that leads to the final state *C* (here we are applying this dynamical description to the problem of abduction). This conformation can be changed only by drastic and catastrophic changes (in the last section these ideas will be exemplified by the historical example of the discovery of Neptune).

To better understand creative cognitive processes and processes of discovery in science, we can integrate this geometrical metaphor with the concepts of *anticipation* and *adumbration*. We want to couple the concept of parameter with adumbration; and the concept of attractor with anticipation.

4.1 Hypotheses anticipation

The philosophical tradition of phenomenology fully recognizes the important role of perceptual and kinesthetic data in the generation of "idealities'" and mental constructs. For example, in phenomenological words, perception is a "structured" *intentional* constitution of the external objects, established by the rule-governed activity of consciousness. The modality of appearing in perception is already markedly structured: it is not that of concrete material things immediately given, but it is mediated by sensible schemata constituted in the temporal continual mutation of adumbrations. So at the level of "presentational perception" of pure lived experiences, only partial aspects (adumbrations [*Abschattungen*]) of the objects are provided. Therefore, a further activity of unification of the different adumbrations to establish they belong to a particular and single object (*noema*) is required. The appearances are the objects as they are intuitively and immediately given (by direct acquaintance) in the constituting multiplicity of the so-called adumbrations, endowed with a morphological character. When we see a - potential, we cannot foretell what it is - spherical form from one perspective, we are adumbrating it.

We have to start dealing with the problem of the treatment of adumbrations. The adumbrative aspects of things are part of the visual field. To manage them a first requirement is related to the need of gluing different fillings-ins of the visual field to construct the temporal continuum of perceptive adumbrations in a global space.

The kinesthetic control of perception is related to the problem of the generation of the objective notion of three-dimensional space, that is, to the phenomenological constitution of a "thing",[9] as a single body unified through the multiplicity of its appearances. The "meaning identity" of a thing is of course related to the continuous flow of adumbrations: given the fact that the incompleteness of adumbrations implies their synthetic consideration in a temporal way, the synthesis is, in this case, *kinetic*, involving eyes, body, and objects. Visual sensations are not sufficient to constitute objective spatiality. Kinesthetic sensations[10] (relative to the movements of the perceiver's own body)[11] are required.

Kinesthetic controls are kinds of *spatial* gluing operators. They are able to compose, in the case of visual field, different partial aspects - identifying them as belonging to the same object, that is constituting an ideal and transcendent "object". They are realized in the pure consciousness and are characterized by an intentionality that demands a temporal lapse of time.

Adumbrations are multiple and infinite, and there is a potential co-givenness of some of them (those potentially related to single objects). Adumbrations, as rough information that has to be further processed, influence the parameters governing the cognitive system, in the sense that they are responsible for its shifts in the state space. They are incomplete and partial so for the complete givenness of an object a temporal process is necessary. *An-*

[9] Cf. also Husserl, 1931 [1913], § 40, p. 129.

[10] On some results of neuroscience that corroborate and improve several phenomenological intuitions cf. Pachoud, 1999, pp. 211-216, Barbaras, 1999, and Petit, 1999.

[11] The ego itself is only constituted thanks to the capabilities of movement and action.

ticipations are the operations necessary to manage adumbrations that have to be performed by objective transcendence. Just because defeasible, anticipations correspond to a kind of non-intuitive intentional expectation. When we see a spherical form from one perspective (as an adumbration), we will assume that it is effectively a sphere, but it could be also a hemisphere (an example already employed by Locke). Anticipations share with visual and manipulative abduction various features: they are highly conjectural and non-monotonic, so wrong anticipations have to be replaced by other plausible ones. Moreover, they constitute an activity of "generate and test" as a kind of "manipulative" cognition: indeed the finding of adumbrations involves kinesthetic controls, sometimes in turn involving manipulations of objects; but the activity of testing anticipations also implies kinesthetic controls and manipulations.

Finally, not all the anticipations are informationally equivalent and work like attractors for privileged individuations of objects: they foretell subsequent *new trends*. In this sense the whole activity is toward "the best anticipation", the one that can display the object in an optimal way. Prototypical adumbrations work like structural-stable systems, in the sense that they can "vary inside some limits", without altering the apprehension of the object. Like in the case of selective abduction, anticipations are able to select possible paths for constituting objects, actualizing them among the many that remain completely tacit. Like in the case of creative abduction, they can construct new ways of aggregating adumbrations, by delineating the constitution of new objects/things. In this case they originate interesting new "attractors" that give rise to new "conceptual" generalizations.

Let us illustrate an astronomical example of anticipation coming from the analysis of the evolution of the cognitive system expressing classical physics: new problems arose after Uranus was accepted to be a planet. Uranus' orbit could not be accurately predicted from Newtonian theory. In fact, by looking at the predicted orbit with a telescope, it was not possible to observe any astronomical body. This was an interesting *anomaly* to be solved. To explain this inconsistency, Adams and Leverrier, in the first half of the nineteenth century, introduced the *ad hoc* hypothesis that this anomaly could be explained by postulating the existence of another still unobserved planet. This is a case of productive *ad hoc* hypothesis guessing. In 1846 Galle decided to point his telescope in the direction indicated by the new hypothesis to effectively determine the existence of the planet. He actually "discovered" Neptune. It was the *decision* to use an external artifact able to "prosthesize" scientists' cognitive skills to produce this scientific "anticipation", concerning an empirical discovery.

Metaphorically we can say that the telescope, as an external tool manipulated by the scientist, "bumped" against the existing attractor accounting for the belief in the orbit of Uranus as predicted by the Newtonian theory. This brought to a catastrophic rearrangement of parameters, that is to the discovery of a new planet and to the development of a new conception of the solar system.

CONCLUSION

We have illustrated the strategic role played by the recent cognitive and epistemological concept of model-based and manipulative abduction. Manipulative abduction has been considered a particular kind of embodied reasoning that exploits external models endowed with delegated cognitive roles and attributes. Abductive manipulations operate on models that are external and the strategy that organizes the manipulations is unknown *a priori*.

We have also said that some aspects of abductive reasoning can be usefully grasped through the perspective of dynamical systems. Creative and selective abduction can be viewed as a kind of process related to the transformations of the *attractors* responsible of the cognitive system behavior. In the context of naturalized phenomenology we have described anticipation and abduction in the light of catastrophic rearrangement of attractors. A perspective that can be further developed for example to treat other interesting aspects of scientific discovery (conceptual change and scientific revolutions).

The ideas elaborated in the described dynamical paradigm are useful to suggest further *naturalized* improvements to the Husserlian phenomenology, by describing the concepts of *adumbration* and *anticipation* in term of attractors and parameters.

REFERENCES

Barbaras, R., (1999). The movement of the living as the originary foundation of perceptual intentionality, in: J. Petitot, F.J. Varela, B. Pachoud, and J.-M. Roy, eds., pp. 525-538.

Boutilier, C. and Becher, V. (1995). Abduction as belief revision, *Artificial intelligence 77,* pp. 43-94.

Bylander, T., Allemang, D., Tanner, M.C., and Josephson, J.R. (1991). The computational complexity of abduction, *Artificial Intelligence 49*, pp. 25-60.

Fodor, J.A., Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis, *Cognition*, 28, pp. 3-71.

Giere, R.N., (1999). Using models to represent reality, L. Magnani, N.J. Nersessian, and P. Thagard, (eds.), pp. 41-57.

Hirsch, M.W. (1984). The dynamical systems approach to differential equations, *Bulletin of the American Mathematical Society*, 11, pp.1-64.

Husserl, E., (1931). *Ideas. General Introduction to Pure Phenomenology* (First book, 1913), translation by W.R. Boyce Gibson, George Allen & Unwin, London and Humanities Press, New York.

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

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press.

Magnani, L. (1999). Creative abduction and hypothesis withdrawal in science, J. Meheus (ed.), *Proceedings of International Congress on Discovery and Creativity (ICDC)*, Dordrecht, Kluwer.

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

Magnani, L., Dossena, R. (2002). Perceiving the infinite and the infinitesimal world: unveiling and optical diagrams and the construction of mathematical concepts*, Foundations of Science*, this issue.

Magnani, L., Nersessian, N.J. (eds.) (2002). *Model-Based Reasoning: Sciences, Technology, Values*. New York, Kluwer Academic/Plenum Publishers.

Magnani, L., Nersessian, N.J., Thagard, P. (eds.) (1999). *Model-Based Reasoning in Scientific Discovery*. New York, Kluwer Academic/Plenum Publishers.

Mill, J.S., (1843). *A System of Logic*. Reprinted in: *The Collected Works of John Stuart Mill*, J.M. Robson, (ed.), London, Routledge and Kegan Paul.

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

Nersessian, N.J. (1999). *Model-based reasoning in conceptual change*, L. Magnani, N.J. Nersessian, and P. Thagard, (eds.), pp. 5-22.

Pachoud, B., (1999). The teleological dimension of perceptual and motor intentionality, in: J. Petitot, F.J. Varela, B. Pachoud, and J.-M. Roy, eds., pp. 196-219.

Peirce, C.S., (1955). Abduction and induction, *Philosophical Writings of Peirce*, J. Buchler, (ed.), New York, Dover, pp. 150-156.

Peirce, C.S. (1931-35, 1958). *Collected Papers* 1-6, C. Hartshorne and P. Weiss, eds.; 7-8, A. Burks, ed., Cambridge, MA, Harvard University Press.

Petit, J.-L., 1999, Constitution by movement: Husserl in the light of recent neurobiological findings, in: J. Petitot, F.J. Varela, B. Pachoud, and J.-M. Roy, eds., pp. 220-244.

Petitot, J., Varela, F., Pachoud, B., Roy, J.M. (eds.) (1999). *Naturalizing Phenomenology*. Stanford, CA, Stanford University Press.

Port, R.F., Van Gelder, T. (eds.) (1995). *Mind as Motion. Exploration in the Dynamics of Cognition*. Cambridge, MA, The MIT Press.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton, NJ, Princeton University Press.

Thom, R. (1980). *Modèles Mathématiques de la Morphogenèse*. Paris, Christian Bourgois (English translation by W.M. Brookes and D. Rand, *Mathematical Models of Morphogenesis*, Chichester, Ellis Horwood).