The Logic of Discovery in the Cyberage

Lorenzo Magnani, Matteo Piazza, and Riccardo Dossena
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, rdossena@pure-reason.unipv.it

Abstract: More than a hundred years ago, the American philosopher C.S. Peirce suggested the idea of *pragmatism* as a logical criterion to analyze what words and concepts express through their practical meaning. Many words have been said on creative processes and reasoning, especially in the case of scientific practices. In fact, philosophers have usually offered a number of ways of construing hypotheses generation, but all aim at demonstrating that the activity of generating hypotheses is paradoxical, illusory or obscure, and then not analyzable. To dismiss this tendency and gain interesting insight about the so-called "logic of discovery" we need to build constructive procedures, which could play a role in moving the problem solving process forward, by implementing them in some actual models. The "computational turn" gave us a new way to understand creative processes in a strictly pragmatic sense. Artificial Intelligence and Cognitive Science tools allow us to test concepts and ideas previously conceived in abstract terms. It is in the perspective of these *actual models* that we find the central role of *abduction* in the explanation of creative reasoning. Creativity and discovery are no more seen as mysterious irrational processes, but, thanks to the constructive accounts, as a complex relationship among different inferential steps (which can also be inductive or deductive), the nature of which can be fruitfully analyzed and identified. We maintain that concepts of *model-based* and *manipulative* abduction are interesting not only in delineating the actual practice of abduction, but also in developing programs computationally adequate in rediscovering, or discovering for the first time, for example, scientific hypotheses or mathematical theorems.

1. PRAGMATISM AS A RULE FOR CLARITY

In the second half of the XIX Century the great American philosopher Charles Sanders Peirce suggested the idea of *pragmatism* as a logical criterion to analyze what words and concepts express through their practical meaning. In *The fixation of belief* (1877) Peirce enumerates the four main methods by means of which it is possible to fix beliefs. Only by means of the method of science, thanks to rigorous research, "we can ascertain by reasoning how things really and truly are; and any man, if he have sufficient experience and he reason enough about it, will be led to the one True conclusion" (Peirce, 1986, p. 255).

Peirce will explain more clearly the public notion of truth and the interpretation of reality as the final purpose of the human inquiry in his subsequent paper, *How to make our ideas clear* (1878). Here Peirce points attention on the notions of "clear idea" and "belief".

"Whoever has looked into a modern treatise on logic of the common sort, will doubtless remember the two distinctions between *clear* and *obscure* conceptions, and between *distinct* and *confused* conceptions" (Peirce, 1986, p. 257). In this paper Peirce is clearly opposing traditional philosophical positions such as those by Descartes and Leibniz, who considered clarity and distinction of ideas only from a merely psychological and analytical perspective:

> It is easy to show that the doctrine that familiar use and abstract distinctness make the perfection of apprehension has its only true place in philosophies which have long been extinct; and it is now time to formulate the method of attaining to a more perfect clearness of thought (Peirce, 1986, p. 258).

Where do we have, then, to look for a criterion of clarity, if philosophy has become too obscure, irrational and confusing, if "for an individual, however, there can be no question that a few clear ideas are worth more than many confused ones"?

"The action of thought is excited by the irritation of doubt, and ceases when belief is attained; so that the production of belief is the sole function of thought" (Peirce, 1986, p. 261). And belief "is something that we are aware of […]; it appeases the irritation of doubt; and, third, it involves the establishment in our nature of a rule of action, or, say for short, a habit" (Peirce, 1986, p. 263). The whole function of thought is to produce habits of action. This leads directly to the *methodological* pragmatic theory of meaning:

> To develop its meaning, we have, therefore, simply to determine what habits it produces, for what a thing means is simply what habits it involves. Now, the identity of a habit depends on how it might lead us to

> act […]. Thus, we come down to what is tangible and conceivably practical, as the root of every real distinction of thought, no matter how subtile it may be; and there is no distinction of meaning so fine as to consist in anything but a possible difference of practice (Peirce, 1986, pp. 265-266).

In this way Peirce sets the equivalence among idea, belief and habit, and can define the rule by means of which we can reach the highest grade of intellectual clearness, pointing out that is impossible to have an idea in our minds which relates to anything but conceived sensible effects of things.

> Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object (Peirce, 1986, p. 266).

This rule founds the pragmatic procedure thanks to which it is possible to fix our ideas.

2. ACTUAL MODELING AND THE PROBLEM OF CREATIVE REASONING

Peirce's conception of clarity contains the idea that to define the meaning of concepts we have to "test", to "subject to" them: the whole conception of some quality lies in its conceivable effects. As he remembers us by the example of the concept of *hardness* "there is absolutely no difference between a hard thing and a soft thing so long as they are not brought to the test" (Peirce, 1986, p. 266). Hence, we can define the "hardness" by looking to those predictable events that occur every time we think to test some thing.

This methodological criterion can be useful to solve the problem of *creative* reasoning, and to describe, in rational terms, some aspects of the delicate question of a "logic of discovery": what do we mean by "creative", and how can be a "creative process" described?

Much has been said on the problem of *creativity* and hypotheses generation. In the history of philosophy there are at least three important ways for designing the role of hypothesis generation, considered in the perspective of problem solving performances. But all aim at demonstrating that the activity of generating hypotheses is paradoxical, either illusory or obscure, implicit, and not analyzable.

Plato's doctrine of *reminiscence* can be looked at from the point of view of an epistemological argument about the paradoxical concept of "problem-solving": in order to solve a problem one must in some sense already know the answer, there is no real generation of hypotheses, only recollection of them. The activity of Kantian *schematism* is implicit too, resulting from

imagination and completely unknowable as regards its ways of working, empty, and devoid of any possibility of being rationally analyzed. It is an activity of tacit knowledge, "an art concealed in the depths of the human soul, whose real modes of activity nature is hardly likely ever to allow us to discover, and to have open to our gaze" (Kant, 1929, A141-B181, p. 183). In his turn Polanyi thinks that if all knowledge is explicit and capable of being clearly stated, then we cannot know a problem or look for its solution; if problems nevertheless exist, and discoveries can be made by solving them, we can know things that we cannot express: consequently, the role of so-called *tacit knowledge* "the intimation of something hidden, which we may yet discover" is central (Polanyi, 1966).

On the other hand philosophers of science in the twentieth century have traditionally distinguished between the logic of discovery and the logic of justification. Most have concluded that no logic of discovery exists and, moreover, that a *rational* model of discovery is impossible. In short, scientific creative reasoning should be irrational and there is no reasoning to hypotheses.

In all these descriptions, the problem is that the definition of concepts like "creativity" and "discovery" is *a priori*. Following Peirce, the definition of concepts of that sort has not usually rested upon any observed facts, at least not in any great degree; even if sometimes these beliefs are in harmony with natural causes. They have been chiefly adopted because their fundamental propositions seemed "agreeable to reason". That is, we find ourselves inclined to believe them.

Usually this frame leads to a proliferating verbosity, in which theories are often incomprehensible and bring to some foresights just by intuition. But a theory which needs intuition to determine what it predicts has a poor explanatory power. It just "makes of inquiry something similar to the development of taste" (Peirce, 1986, p. 254).

A suggestion that can help to solve the enigma of discovery and creativity comes from the "computational turn", developed in the last years. Recent computational research in the field of cognitive science makes use of tools able to give up those puzzling "speculative" problems, or, at least, to redefine them in a strict pragmatical sense. In fact, modern tools of logic, artificial intelligence and computational philosophy permit to construct actual models of the studied processes. It is an interesting constructive rational alternative that, by disregarding the most abstract level of analysis, can offer clear and testable architectures of creative processes.

3. FROM IRRATIONALITY TO ABDUCTION

Inside the computational frame, a new paradigm rises by emphasizing the significance of the concept of *abduction*. It aims at illustrating the problem-solving process and at proposing a unified and rational epistemological model of scientific discovery, diagnostic reasoning, and other kinds of creative reasoning.

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")[1] and 2) abduction considered as inference "to the best explanation", which also evaluates hypotheses.

Theoretical abduction[2] 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*[3] 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 (more details on this kind of reasoning are given in the following section).

The type of inference called abduction was studied by Aristotelian syllogistics, as a form of α¹παγωγη¿, and later on by mediaeval reworkers of syllogism. A hundred years ago, Peirce interpreted abduction essentially as an "inferential" *creative process* of generating a new hypothesis, as the only

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

[2] 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.

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

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*, that is in terms of the "classical" logic:

> If a patient is affected by a pneumonia, his/her level of white blood cells is increased.
> John is affected by a pneumonia.
> John's level of white blood cells is increased[4].
> (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 (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

[4] 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.

generalization, like the piece of hematologic knowledge represented by (1)[5]. Starting from knowing – selecting – (1) and "observing" (3) we can infer (2) by performing a selective abduction[6]. The abductive inference rule corresponds to the well-known fallacy called affirming the consequent (simplified to the propositional case)

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

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.

Medical diagnosis offers the concrete evidence of an "actual" model of abduction. Researchers involved in building diagnostic "KBSs" (knowledge-based systems) in medical domains needed an epistemological clarification: to this aim the concept of abduction is fundamental. In these cases the abductive selection and evaluation of hypotheses is central and clearly shown as concretely "constrained" (Ramoni et al., 1992).

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 the selective side of abduction (diagnostic reasoning) and relates to the idea of preserving *consistency*.

Formal and informal models of abduction are "in place". To summarize, in the perspective of classical logic abduction is a fallacy; outside of the classical approach, many logical models (that is deductive models) have

[5] 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].

[6] 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.

been built, mainly concerning selective aspects of abduction and its non-monotonic character. We would like to stress that they constitute "models" of abductive processes which do not have special epistemological privileges. Among other models, computational, psychological, anthropological, philosophical, etc., they participate of the nature of abduction by constituting *actual* practices. The same for the informal logic models developed to account for reasoning involved in actual arguments related to abductive processes. It is important to remember that many ways of performing "selective" abductive inferences are in place and ready to be used, but always contingent and particular.

By the way, exclusively considering the sentential view of abduction cited above 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.

Hence abduction has to be seen in a *complex relationship* among many inferential steps, deductive, inductive, etc., so that its mysterious aura disappears. This is a central point. When faced to the problem of affording new perspectives on abduction, Peirce and AI tradition immediately suggest that we have to adopt a dynamical and interdisciplinary attitude. Some research in the area of artificial intelligence (for example diagnostic reasoning and machine discovery) has shown that methods for discovery can be found that are computationally adequate for rediscovering - or discovering for the first time - empirical or theoretical laws and theorems. Moreover, the study of visual, spatial, analogical, and temporal reasoning, in epistemology, logic, and AI has demonstrated that there are many ways of performing intelligent and creative reasoning that cannot be described with only the help of classical logic. However, non-standard logic has shown how we can provide rigorous formal models of many kinds of abductive reasoning such as the ones involved in certain defeasible and uncertain inferences (see above).

Finally, we have cited above the IbE (inference to the best explanation), one of the most appealing "abstract" aspects of the abductive reasoning architectures (for example in science, where new abduced hypotheses and theories are considered epistemically better than the previous ones). We have to stress it is always performed in a framework of local and *actual* constraints and criteria. Of course these "local" inferences to the best explanation still provide some sort of "general" ways of reasoning shared by different settings. Anyway, these reasoning processes cannot be considered uni-

versal and/or "epochally" constant (for example, selecting the best scientific hypothesis in an encyclopedia of available hypotheses in medical diagnosis is different from judging the best moral hypothesis to adopt a possible subsequent action; modern science conceptual changes in terms of best explanations obey to criteria very different from the ones used in ancient science). We are always concerned with actual and local criteria also when we have to build abductive computational programs able to perform IbE. From the strict philosophical perspective the problem can be only depicted in an abstract and general sense.

3.1 Model-based abduction and its external dimension

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 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" (Peirce, 1931-35, 1958, 5.283), and, in Kantian words, all synthetic forms of cognition. That is, a considerable part of the thinking activity is model-based. Model-based reasoning acquires its peculiar creative relevance when embedded in abductive processes, so that we can individuate a *model-based abduction*.

Following Nersessian (Nersessian, 1995, 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 (see also Johnson-Laird, 1983). Surely it is the actual practice of abduction that helps to see the roots of abduction in the manipulation of images and objects. Obvious examples of model-based reasoning are constructing and manipulating visual representations, thought experiment or analogical reasoning.

Manipulative abduction (Magnani, 2001) - contrasted to theoretical abduction as a particular form of model-based 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.

Peirce considers inferential any cognitive activity whatever, not only conscious abstract thought; he also includes perceptual knowledge and subconscious cognitive activity. 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" (1931-35, 1958, 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 that a creative form of thinking through doing (manipulative abduction)[7] is fundamental as much as the theoretical one. The whole activity of manipulation is devoted to building various external *epistemic mediators*[8] that function as an enormous new source of information and knowledge: a kind of redistribution of the epistemic and cognitive effort to manage objects and information that cannot be immediately represented or found internally. This external representation can modify, simplify, and improve the kind of computation that a human agent uses to reason about a problem.

4. CONCLUSION

We have seen that, to solve the problem of the so-called "logic of discovery", we have to clarify what we are looking for, and the meaning of concepts like *creative* and *discovery*. Following Peircian ideas, we can see the recent computational tools and modeling activities as a useful support to philosophy in a strictly pragmatical sense. We can search and implement actual *rational* models of creative reasoning and scientific discovery.

In this intellectual framework a new paradigm aimed at unifying the different perspectives is played by the fundamental concept of *abduction*. Many "working" abductive processes can be found and studied that are rational,

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

[8] This expression, suggested by Magnani (2001), is derived from the cognitive anthropologist Hutchins (1995), 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 "primitive" settings.

unambiguous, and perfectly communicable. We maintain that the concepts of *model-based* and *manipulative* abduction are crucial not only in delineating the actual practice of abduction, but also in the development of programs computationally adequate in rediscovering, or discovering for the first time scientific hypotheses or mathematical theorems.

REFERENCES

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.

Cornuéjols, A., Tiberghien A., and Collet, G. (2000). A new mechanism for transfer between conceptual domains in scientific discovery and education, *Foundations of Science*, *5(2)*, Special Issue on "Model-Based Reasoning in Science: Learning and Discovery", L. Magnani, N.J. Nersessian, and P. Thagard eds., pp. 129-155.

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

Kant, I., 1929, *Critique of Pure Reason*, translated by N. Kemp Smith, MacMillan, London, reprint 1998; originally published 1787.

Magnani, L. (2001). *Abduction, Reason, and Science. Processes of Discovery and Explanation*. 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, *Model-based reasoning in conceptual change*, L. Magnani, N.J. Nersessian, and P. Thagard, (eds.), pp. 5-22.

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.

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

Peirce, C.S. (1986). *Writings of Charles Sanders Peirce, A Chronological Edition*. Bloomington, IN, Indiana University Press, Vol. 3 (1872-1878).

Polanyi, M. (1966). *The Tacit Dimension*, Garden City, New York, Doubleday and Co.

Ramoni, M. Stefanelli, M., Magnani, L., and Barosi, G. (1992), An epistemological framework for medical knowledge-based systems, *IEEE Transactions on Systems, Man, and Cybernetics* 22(6):1361-1375.