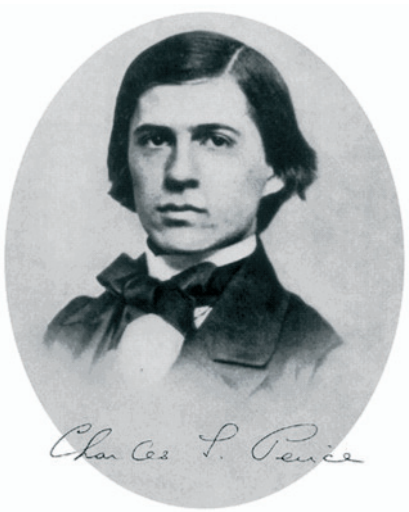
Titulo: Abduction and the Process of Scientific Discovery
Edição: Olga Pombo e Alexander Gerner
Centro de Filosofia das Ciências da Universidade de Lisboa
Faculdade de Ciências da Universidade de Lisboa
Campo Grande, Edifício C1, 3º Piso, Sala 1.3.35
1749-016 Lisboa
E-mail: cfcul@fc.ul.pt
Data: Setembro de 2007
ISBN:
Depósito legal:
© Copyright do Autor

Este livro foi realizado no âmbito do Centro de Filosofia das Ciências da Universidade de Lisboa (POCTI-ISFL-20-678), com o apoio e a aprovação da FCT e do POCTI e comparticipado pelo fundo comunitário FEDER.

Abduction and the Process of Scientific Discovery

Edited by Olga Pombo and Alexander Gerner



Papers delivered at the International Meeting "Abduction and the Process of Scientific Discovery" promoted by the Center for Philosophy of Science, University of Lisbon (CFCUL) and held at the Science Museum, University of Lisbon, 4th-6th May 2006.

Colecção Documenta

Centro de Filosofia das Ciências da Universidade de Lisboa

Table of Contents

Preface 7

Acknowledgments 10

1. Peirce and the Concept of Abduction

Abduction, Belief-Revision and Non-Normality 13
Shahid Rahman

What is Abductive Method? 27
Pasi Pohjola

Abductive Logic of Discovery with Distributed Means 47
Sami Paavola

2. Processes of Scientific Discovery

Prospective Updating of Theories with Preferences 65
Luís Moniz Pereira, Pierangelo Dell'Acqua and Gonçalo Lopes

Adaptive Logics for Abduction and the Explication of Explanation-Seeking Processes 97
Joke Meheus

Content Guidance in Formal Problem Solving Processes 121
Diderik Batens

3. Case Studies on Abduction

How can the Example of Kepler's Great Scientific Discovery Illuminate the Problematical Nature of Abduction? 159
Ana Paula Silva

Abduction as a Strategy for Concept Formation in Mathematics: Cardano Postulating a Negative 179
Albrecht Heeffer

Is the Major Conceptual Moment of the Orthodox Quantum Mechanics a Moment of Abduction? 195
João Luís Cordovil

Abductive Reasoning, Theoretical Preduction, and the Physical Way of Dealing Fallibly with Nature 199
Andrés Rivadulla

4. Abductive Strategies, Applications & Perspectives

Seeing Problems, Seeing Solutions. Abduction and Diagrammatic Reasoning in a Theory of Learning and Creativity 213
Michael Hoffmann

Alterity and Dialogism in Abductive Processes of Scientific Discovery 237
Augusto Ponzio

Abduction and Metaphor 268
José Augusto Mourão

Abduction, Fallacies and Rationality in Agent-Based Reasoning 283
Lorenzo Magnani and Elia Belli

Abductive Issues in the Proof of Pragmaticism 303
Ahti-Veikko Pietarinen

Abductions Dark Ways 321
Willem F. G. Haselager

Clinical Reason in Medicine: The Role of Abductive Inference 335
João Sampaio

Scientific Discovery and Social Symptomatology: Abduction, Medical Semeiotics and Semioethics 341
Susan Petrilli

Neural Processes, External Semiotic Anchors, Hybrid Representations 365
Lorenzo Magnani

Preface

Abduction is the process of forming an explanatory hypothesis.
It is the only logical operation which introduces any new idea
(Peirce, 1903, CP 5.171)

Charles Saunders Peirce is the well known founder of the concept of "abduction" and related theory of scientific discovery. An idea which has proved to be of vast applicability in very different areas such as Philosophy of Science, Philosophy of Mind, Logics, Cognitive Science, Semiotics, Artificial Intelligence or Medical Diagnosis. Aiming to contribute to the development of this upcoming field of research, the *Center for Philosophy of Science of the University of Lisbon* (CFCUL) gave rise, in 2005, to a Peirce-Study-Group intending to pursue interdisciplinary analysis on Peirce' s lifelong studies on abduction and scientific discovery. This research group developed a project which was submitted to the *European Science Foundation* (ESF), as an exploratory workshop. The idea was to pave the way for a desirable methodology of heuristics going beyond the classical inductive-deductive scheme of scientific reasoning.

The support of the ESF was not granted. Nonetheless, the CFCUL decided to organize an international meeting entitled *Abduction and the Process of Scientific Discovery*. The meeting's general endeavor was to clarify up-to-date terminologies and concepts of abduction; to discuss derived abdutive methodologies and epistemic strategies in scientific discovery, both today and in a historical perspective; to introduce examples of actual applications and possible future perspectives linked to the study of abduction and the process of scientific discovery.

The meeting took place in Lisbon, from 4th to 6th May 2006 and we had the great honor of receiving all the invited participants who desired to come even if unfortunately we could not cover all the costs involved for the participants. This means that the scientific spirit to meet on such an important topic prevailed and made this meeting possible as a significant continuation of the ongoing debate on the legacy of Charles Saunders Peirce for the Philosophy of Science and various related research fields.

We deeply regret that some of the papers given at the meeting were not sent for publication: Albert Schirrmeister, "Johannes Keplers Dreams and Abductive Knowledge", Dina Mendonça and Isabel Serra, "Abduction and Emotional Processes - a Taxonomy of Emotional Processes in Scientific Discovery" and Nuno Nabais, "Abduction and Categories". On the other hand we are glad to be able to additionally include in this volume two papers which regrettably could not be presented at the conference (Susan Petrilli, "Scientific Discovery and Social Symptomatology: Abduction, Medical Semeiotics and Semioethics"), and Augusto Ponzio ("Alterity and Dialogism in Abductive Processes of Scientific Discovery"), but which hereby are made accessible to the public.

*

The present volume is mostly organized accordingly to the four topics of the Lisbon's International Meeting *"Abduction and the Process of Scientific Discovery"*.

The first chapter is devoted to *Peirce and the Concept of Abduction*. Here issues linked to the conceptual, terminological and epistemological determination of abduction and abductive method starting from Peirce are focused. This section includes papers by Shahid Rahman ("Abduction, Belief-Revision and Non-Normality"), Pasi Pohjola ("What is Abductive Method?"), Alexander Gerner ("Abduction and Epistemic Attention") and Sami Paavola ("Abductive Logic of Discovery with Distributed Means").

In the second chapter the topic is *Processes of Scientific Discovery.* It presents the papers related to the computational and logical approach towards abduction, namely those of Luís Moniz Pereira, Pierangello Dell'Acqua and Gonçalo Lopes ("Prospective Updating of Theories with Preferences"), Joke Meheus ("Adaptive Logics for Abduction and the Explication of Explanation-Seeking Processes"), Diderik Batens ("Content Guidance in Formal Problem Solving Processes").

The third chapter, with the title *Case Studies on Abduction,* includes the papers given by Ana Paula Silva ("How can the Example of Kepler's Great Scientific Discovery Illuminate the Problematical Nature of Abduction?"), Albrecht Heeffer ("Abduction as a Strategy for Concept Formation in Mathematics: Cardano Postulating a Negative"), João Luís Cordovil ("Is the Major Conceptual Moment of the Orthodox Quantum Mechanics a Moment of Abduction?") and Andrés Rivadulla ("Abductive Reasoning, Theoretical Preduction, and the Physical Way of Dealing Fallibly with Nature").

Finally, the fourth chapter, *Abductive Strategies, Applications & Perspectives,* contains the papers delivered by Michael Hoffmann ("Seeing Problems, Seeing Solutions. Abduction and Diagrammatic Reasoning in a Theory of Learning and Creativity"), Augusto Ponzio ("Alterity and Dialogism in Abductive Processes of Scientific Discovery"), José Augusto Mourão ("Abduction and Metaphor"), Lorenzo Magnani and Elia Belli ("Abduction, Fallacies and Rationality in Agent-Based Reasoning"), Ahti-Veikko Pietarinen ("Abductive Issues in the Proof of Pragmaticism"), Willem F. G. Haselager ("Abductions Dark Ways"), João Sampaio ("Clinical Reason in Medicine: The Role of Abductive Inference"), Susan Petrilli ("Scientific Discovery and Social Symptomatology: Abduction, Medical Semeiotics and Semioethics"), and Lorenzo Magnani ("Neural Processes, External Semiotic Anchors, Hybrid Representations").

Olga Pombo and Alexander Gerner

Acknowledgments

The CFCUL thanks all the participants who made this meeting a great and successful encounter on such an important topic as abduction and scientific discovery. Special thanks are due to Prof. Shahid Rahmann (University of Lille III/France), without whom the meeting would not have had so lively discussions, and with whom the CFCUL continues to work in various initiatives. The CFCUL moreover thanks the Portuguese Foundation for Science and Technology (FCT) and the Luso-American Foundation for Development (FLAD) for the financial support given to the conference and to the edition of the proceedings here presented. Also we would like to thank FLAD for making possible the stay of Prof. Michael Hoffmann from the Georgia Institute of Technology, USA, who, besides his contribution to the meeting, could moreover present a conference in the *Permanent Seminary of Philosophy of Science* regularly promoted by the CFCUL.

We want yet to express our thanks to João Luís Cordovil, Ana Paula Silva, Ricardo Coelho, João Sampaio, Golgona Anghel and Eduardo Pellejero and also the former secretary of the CFCUL Elisabete Fonseca who all helped very much with their enthusiasm for the successful result of this meeting in a pleasant and dynamic working atmosphere.

Finally, we would like to thank the Science Museum of the University of Lisbon which generously gave us the possibility of presenting this meeting at the beautiful historic buildings of the old Faculty of Sciences thus offering the additional *beauty*, *dignity* and *inspiration* which - one should not forget to mention - are necessary elements when dealing with *scientific discovery* and the *origin of great innovative ideas.*

1

PEIRCE AND THE CONCEPT OF ABDUCTION

Abduction, Belief-Revision and Non-Normality

Shahid Rahman
Université de Lille 3
U.F.R. de Philosophie
shahid.rahman@univ-lille3.fr

Introduction

Let me start by taking up the quote chosen by the organisers of this excellent conference:

> Abduction is the process of forming an explanatory hypothesis.
> It is the only logical operation which introduces any new idea.

This quote already contains most of the challenges of Peirce's notion of abduction. In an earlier very similar quote (2.777) in a chapter on *ampliative* reasoning and presumption which he proposes there to call *abduction* Peirce claims that abduction processes are synthetic.

The points I want to stress of this quote are

- the element of surprise
- the logical feature, and
- the process or operative conception

which underlie Peirce's notion of abduction.

It has been remarked by Robert Burch that as Peirce came to extend and generalize his notion of abduction, abduction became defined as inference to and provisional acceptance of an explanatory hypothesis for the purposes of testing it. Abduction is not always inference to the best explanation, but it is always inference to some explanation or at least to something that clarifies

some information that has previously been "surprising," in the sense that we would not have routinely expected it, given our then-current state of knowledge. Abduction triggers revision of our previous beliefs, but this revision should be carried over in a rationally plausible way.

In the nineteen-sixties, when philosophers of science, influenced by the Aristotelian view on explanation, were trying to build a logical notion of explanation, abduction arose as challenging the very conceptual basis of the project. Abduction seems to talk about the exceptional and the surprising; but this is in flagrant conflict with the logic the philosophers of science were trying to apply. The logic they were thinking of is about the general and is quite often seen as condemned to be unsurprising. Moreover it is seen as having an analytic rather than a synthetic character.

I think that this tension between the pairs *general-analytic* and *exceptional-synthetic* radically changed the whole deductive-nomological project. Even if one accepts the idea that explanations are deductive arguments, in the sense of the standard classical logic, it seems an almost impossible task to draw a viable distinction between those arguments that do qualify and those that do not qualify as explanations.

Many philosophers took this difficulty as a reason to abandon once and for all the very idea of a deductive or logical approach to explanation and abduction and replaced it with sociological and psychological frameworks.

It is interesting that within the community of analytical philosophers Sylvain Bromberger rejected already in 1966 Hempel and Oppenheim's covering-law concept, that identifies explanation, or correct answers to why-questions, with logical subsumption under true lawful generalizations. Bromberger's basic idea is that why-questions arise when one believes that the answer must be a departure from a general rule, and thinks that the conditions under which departures from the general rule occur can be generalized.

More recently, Jaakko Hintikka and Ilpo Halonen initiated a thorough and devastating attack on the subsumption theory. Their approach shares with Bromberger the view that explanation as used by scientists is not principally understood as subsuming facts into an already known law, but discov-

ering the relevant law. Their point is that explanation can be conciliated with abduction if we centre our attention in the process of explanation.

Let me formulate the distinctions Bromberger and Hintikka are dealing with in the following way: when we say that an epistemic agent does not know the answer to a why question, we might mean two different things:

1. We might think that there is a right answer to our question somewhere but we do not know how to arrive at it. Let me call this situation: *Not knowing about actions*. Notice that in this case the epistemic agent might include situations in which he actually knows the right answer but he does not know how to arrive at it (think for example of the result of an equation). Bromberger calls such situations *b-predicament situations*.

2. We might think that there is a right answer somewhere but our present state of knowledge is not sufficient to prefer some answers rather than others. Let me call this situation: *Not knowing about propositions*. Notice that this situation includes cases of lack of propositional omniscience. Bromberger calls these situations *p-predicament situations.*

Hintikka's remarkable view, which does apply to contexts beyond explanation and abduction, is that both types of lack of knowledge are logical issues and could be handled within a logical framework.

The aim of my talk has an exploratory character, intended to motivate thoughts on combining Hintikka's view on explanation with two competing paradigms, namely; the belief revision paradigm, which keeps track of the logical process of revising and updating surprising and unsurprising information; with Batens and his collaborators approach on dynamic proof theory. In relation to the latter my contribution restrict itself to suggesting where to apply their so-called *adaptive-logic* methods.

Towards a dynamic proof theory for Bonanno's Multimodal Revision-Logic $\mathcal{R}$

When did you last change your mind? I hope this was not so long ago. Belief revision is about the logic of this revising process and was triggered by the remarkable Alchourrón/Gärdenfors/Makinson paper of 1986 known as the *AGM*.

In a paper published last year in *Synthese* (vol. 147, N2, 2005, 193-228) Giacomo Bonanno develops what he calls *A simple modal logic for belief revision*. In this excellent paper the author presents a straightforward multimodal logic which combines the notions of static belief based on the seminal work of Hintikka with the AGM approach.

Moreover the paper connects the epistemic and the belief revision tradition with the Bayesian approach used to model change of beliefs in the context of economics and game theory. Here again we meet more of Peirce's basic notions in philosophy of science: rationally plausible belief revision and economics.

The core of this contribution is to explore some further developments of Bonanno's proposal in the context of AGM and beyond. More precisely,

- I will show how to produce a proof theory for the semantics of $\mathcal{R}$ in the spirit of Peirce : I will use dialogic as a means to build up a correct and sound proof theory.
- I will mention the principal moves leading to an extension of $\mathcal{R}$ able to eliminate cases of logical omniscience still present in Bonanno's approach.
- I will motivate a further extension of $\mathcal{R}$ for the case of inconsistent beliefs. My suggestion is that we could apply here Batens' dynamic proof theory to the dialogics of $\mathcal{R}$.

The revision logic $\mathcal{R}$

The idea is to introduce three epistemic modal operators. A *belief-operator* related to some stock of initial information, an *information-operator* which could trigger an *updating of the initial stock*, and a *revising belief-operator* which considers the consequences of the new information. The intended interpretations of these operators are thus:

1) $B_0\varphi$ initially (at time 0) the individual believes that φ
2) I,φ (between time 0 and time 1) the individual is informed that φ
3) $B_1\varphi$ at time 1 (after revising his beliefs in light of the information received) the individual believes that φ.

Each of these operators will be defined with the help of corresponding accessibility relations to the set of worlds defined in the model.

The aim of the logic $\mathcal{R}$ is to capture the Bayesian approach to belief revision in a modal framework. This approach embodies a conservative principle for belief revision, according to which "*When changing beliefs in response to new evidence, you should continue to believe as many of the old beliefs as possible*".

The conservative principle requires that, if the epistemic agent considers a state possible and the new information does not exclude this state, then he should continue to consider it possible.

The point is that changes should only be allowed if they are prudent.

Let us come to the technicalities.

On the syntactic side we have the standard modal extension of the language of classical propositional logic, which in our case amounts to an addition to the three epistemic operators mentioned above. In regard to the information operator there is some choice here. One might want to restrict this operator to facts disallowing belief-operators in the scope of an information-operator. Without this restriction, we could have the following formula defined as well-formed: $B_0\varphi \wedge I\neg B_0\varphi$ which says that the epistemic agent initially believes φ and is later informed that he did not believe it after all. Such undesirable situations could be also eliminated with the help of suitable axioms rather than changing the notion of a well-formed formula. We will assume one choice, whatever it is, because the choice does not affect the point of $\mathcal{R}$.

DEFINITION : Frame, Model, Truth

A frame for $\mathcal{R}$ is $\langle \mathbf{W}, \mathbf{R}^{B0}, \mathbf{R}^{I}, \mathbf{R}^{B1} \rangle$ for modal propositional logic consists of

- a non empty set **W** of possible worlds (states or scenarios: like temporal states, states of information etc.)
- binary relations $\mathbf{R}^{B0}$, $\mathbf{R}^{I}$, $\mathbf{R}^{B1}$ on **W** called in standard modal logic *accessibility relations*. In the epistemic context the relations that correspond to the belief operators determine the set of states (or worlds) that, the epistemic agent considers possible at a given state. The

relation $\mathbf{R}^{I}$ determine the set of states of information compatible with the information received. More explicitly:

1) $w_i \mathbf{R}^{B0} w_j$ at state w_j the epistemic agent initially (at time 0) considers w_j possible
2) $w_i \mathbf{R}^{I} w_j$ (between time 0 and time 1) the epistemic agent is informed that φ
3) $w_i \mathbf{R}^{B1} w_j$ state w_j the epistemic agent at time 1 (in the light of the information received) considers w_j possible

Let the expression

$\mathbf{R}^{B0}\{w_i\}$

denote the set of states that, initially, the individual considers possible at state w_i. That is,

all $w_j \in \mathbf{W}$ such that $w_i R w_i$

Define $\mathbf{R}^{I}\{w_i\}$ and $\mathbf{R}^{B1}\{w_i\}$ similarly.

Let $[\![\varphi]\!]$ denote *the truth set of* φ; that is, the set *all* worlds (states) of the model where φ is true.

A model for $\mathcal{R}$ results from adding to a frame a valuation function v which assigns a truth value $v(p)$ to each propositional letter of the propositional language in each world $w_i \in \mathbf{W}$:

$<\mathbf{W}, \mathbf{R}^{B0}, \mathbf{R}^{I}, \mathbf{R}^{B1}, v>$

Now, to the main event, the notions of truth in a model and validity:

DEFINITION: If **M** is a model for $\mathcal{R}$ $<\mathbf{W}, \mathbf{R}^{B0}, \mathbf{R}^{I}, \mathbf{R}^{B1}, v>$

, then $v_{Mwi}(\varphi)$ – i.e. the truth value of φ in w_i given M, is defined in the following way:

$v_{Mwi}(p) = v_{Mwi}(p)$, for all propositional letters p
$v_{Mwi}(\neg\varphi) = 1$ iff $v_{Mwi}(\varphi) = 0$
$v_{Mwi}(\varphi \rightarrow \psi) = 1$ iff $v_{Mwi}(\varphi) = 0$ or $v_{Mw}(\psi) = 1$
$v_{Mwi}(\varphi \vee \psi) = 1$ iff $v_{Mwi}(\varphi) = 1$ or $v_{Mwi}(\psi) = 1$
$v_{Mwi}(\varphi \wedge \psi) = 1$ iff $v_{Mwi}(\varphi) = 1$ and $v_{Mwi}(\psi) = 1$
$v_{Mwi}(B_0\varphi) = 1$ iff for at all $w_j \in \mathbf{W}$ such that $w_i \mathbf{R}^{B0} w_j$: $v_{Mwj}(\varphi) = 1$

$v_{Mwi}(\boldsymbol{B_1}\varphi)=1$ iff for at all $w_j \in \mathbf{W}$ such that $w_i \mathbf{R}^{B1} w_j$: $v_{Mwj}(\varphi)= 1$
$v_{Mwi}(\boldsymbol{I}\varphi)=1$ iff for at all $w_j \in \mathbf{W}$ such that $w_i \mathbf{R}^{I} w_j$: $v_{Mwj}(\varphi)= 1$
and there is no other world in W where φ is true.

Note that while the truth conditions for both Belief-operators are the standard ones, the truth condition is rather unusual. It amounts to the requirement that a proposition under the scope of this operator is true iff the set of accessible worlds and the set of worlds where that proposition is true is the **same** set. That is
$v_{Mwi}(\boldsymbol{I}\varphi)=1$ iff $\mathbf{R}^{I}\{w_i\}= \Downarrow\varphi\Diamond$
We will discuss this feature further on, but let me here only make the point that this definition places the information-operator in the field of the so-called non-normal modal operators.

The definitions of modal and frame validity are the standard ones:

DEFINITION : $\mathcal{R}$-Valid

We say that the model $<\mathbf{W}, \mathbf{R}^{B0}, \mathbf{R}^{I}, \mathbf{R}^{B1}, v>$ is *based* on the frame $<\mathbf{W}, \mathbf{R}^{B0}, \mathbf{R}^{I}, \mathbf{R}^{B1}>$

- A formula φ is valid in a model <W, RB0, RI, RB1, v > if it is true at every world of W.
- A formula φ is *valid in a frame* if it is true at every world of every model based on that frame. For short: a formula φ is *valid in a frame* if it is valid in every model based on that frame.

In fact the logic $\mathcal{R}$ implements the Bayesian conservative principle. But as already mentioned this implementation will be performed in a qualitative frame. Modal logic is understood here in the sense of MacColl, who conceived modality as the qualitative version of probability.

In order to achieve this aim Bonanno adds the following three axioms in the multimodal framework described above

QUALIFIED ACCEPTANCE: $(\boldsymbol{I}\varphi \wedge \neg \boldsymbol{B}_0 \neg\varphi) \rightarrow \boldsymbol{B}_1\varphi$

The axiom says that if the epistemic agent is informed that φ and he considered φ already possible then he continues to accept φ in his revised beliefs.

PERSISTENCE: $(I\varphi \wedge \neg B_0\neg\varphi)\rightarrow (B_0\psi\rightarrow B_1\psi)$

The axiom says that if the epistemic agent receives *non-surprising* information (i.e., information that does not contradict his initial beliefs) then he continues to believe everything that he believed before.

MINIMALITY: $((I\varphi \wedge B_1\psi)\rightarrow B_0(\varphi\rightarrow\psi)$

The third axiom says that beliefs should be revised in a logically minimal way, in the sense that no new beliefs should be added unless they are implied by the old beliefs and the information received.

Bonanno proves that these axioms characterize what he calls the *Qualitative Bayes Rule* (QBR). The QBR says that if at a state the information received is consistent with the states considered at the initial beliefs, then the states considered possible according to the revised beliefs are precisely those states:

For all worlds of the model,
if $\mathbf{R}^I\{w_i\} \cap \mathbf{R}^{B0}\{w_i\} =$ then $\mathbf{R}^{B1}\{w_i\} = \mathbf{R}^I\{w_i\} \cap \mathbf{R}^{B0}\{w_i\}$

This brings us to one of the main occupations of modal logicians, which is laying bare the relations between the validity of the formulae and the properties of frames. Indeed, the validity of the three axioms of $\mathcal{R}$ characterize those frames meeting the Qualitative Bayes Rule, in the sense that they are valid *iff* the conditions described the rule hold.

We will tackle the question of characterizing the logic $\mathcal{R}$ by providing a new proof system for this logic by means of ***dialogic***.

Dialogic for $\mathcal{R}$

Modal dialogic is a systematic account of an explicit notion of context, in the sense that the latter is introduced by an explicit label. Modal moves are hence dialogical expressions with a supplementary label, indicating the context in which the move has been made. The Belief operators B_n, where **n** is **0** or **1**, are then defined in the following way:

	Attack	*Defence*
X $\boldsymbol{B}_n A$ I (*BA* has been stated by player X at context i)	Y ? $_{j-}\boldsymbol{B}_n$ i (at the context i the challenger Y attacks by choosing a dialogically context j $\boldsymbol{B}_n$-accessible)	X *A* j (the defender claims that *A* holds at j)
X $\neg\boldsymbol{B}\neg A$ I ($\neg\boldsymbol{B}\neg A$ has been stated by player X at context i)	Y? i (at the context i the challenger Y attacks asking X to choose a j $\boldsymbol{B}_n$-accessible where *A* holds)	X*A* j (the defender chooses the context j and claims that *A* holds there)

The interesting case is the one of the non-standard operator *I*. The challenger can choose here between the two attacks. Namely the standard attack, and the non-standard one. The latter amounts to challenge the claim of the defender that there is no other world in the model, *accessible or not*, where the formula at stake is true. In fact; the challenger will propose a non-accessible world where the formula is true too:

	Attack	*Defence*
X $\boldsymbol{I}A$ i ($\boldsymbol{I}A$ has been stated by player X at context i)	**Standard attack** Y ? $_{j-}\boldsymbol{I}$i (at the context i the challenger Y attacks by choosing a dialogically context j $\boldsymbol{I}_n$-accessible) **Non-standard attack** Y *A* j (at the context i the challenger Y attacks by choosing a dialogically context j NOT $\boldsymbol{B}_n$-accessible and claims that *A* holds there) **The challenger chooses between both attacks**	X *A* j (the defender claims that *A* holds at j) No defence. Only counterattack is possible

In modal dialogic the frame conditions are implemented as special structural rules which allow the Proponent to increase his choice possibilities while challenging a necessity operator or defending a possibility operator. In general it holds that:

MODAL FORMAL RULE

At label **i** the Proponent may choose a label **j** such that **i**R^D**j** iff **j** has been introduced by the Opponent before or this choice has been allowed by the appropriate modal structural rule.

For the logic $\mathcal{R}$ we will have the following version of the Bayesian Qualitative Rule:

DIALOGICAL BQR

At i and **given** i.j-$\boldsymbol{B}_1$, the Proponent can choose this accessibility to challenge a $\boldsymbol{B}_0$-operator and/or $\boldsymbol{I}$-operator.

If we apply this to the proofs of the axioms this will yield

The Proponent can pose an atomic formula at j-$\boldsymbol{B}_1$, if this atomic formula has been posed before by the Opponent at j-$\boldsymbol{B}_0$ and/or j-$\boldsymbol{I}$

Let us prove then the Qualified Acceptance Axiom

$(\boldsymbol{I}p \wedge \neg \boldsymbol{B}_0 \neg p) \rightarrow \boldsymbol{B}_1 p$

	O			P			
				$(\boldsymbol{I}p \wedge \neg \boldsymbol{B}_0 \neg A) \rightarrow \boldsymbol{B}_1 p$		(0)	1
1	(1)	$(\boldsymbol{I}p \wedge \neg \boldsymbol{B}_0 \neg p)$	0		$\boldsymbol{B}_1 p$	(2)	1
1	(3)	? $\boldsymbol{B}_0$-1.1	2		p	(12)	$\boldsymbol{B}_1$-1.1
1	(5)	$\boldsymbol{I}p$		1	?∧1	(4)	1
1	(7)	$\neg \boldsymbol{B}_0 \neg p$		1	?∧2	(6)	1
$\boldsymbol{B}_0$ - 1 . 1	(9)	p		3	? $\boldsymbol{B}_0$	(8)	1.
$\boldsymbol{I}$-1.1	(11)	p		7	? $\boldsymbol{I}$-1.1	(10	1.

P WINS

The Proponent cannot respond immediately to the challenge on the second move. He has to wait until the Opponent has conceded p at the world demanded. To achieve this he has to force the Opponent to concede that precisely at this world p is true and that world is $\boldsymbol{B}_0$ ***and*** $\boldsymbol{I}$-accessible. Actually, in principle the Opponent could have chosen another world than $\boldsymbol{B}_0$-1.1. But

then the Proponent wins by a standard attack on the information operator. Indeed, if the Opponent concedes that there is a world **j** not accessible from 1 where *p* is true, then the Proponent can challenge the information operator claiming that the formula ***I***p does not hold.

Now, certainly for the sake of this sole axiom the attack on the information operator is sufficient and the move 6 is redundant.

Beyond $\mathcal{R}$

Non-normality and omniscience

The interesting point of this logic is that the information operator behaves like a non-normal operator. That is, as an operator where necessitation does not hold. A proposition can thus be valid without being necessarily so. This connects with the well-known debate on eternal truths between Descartes and Leibniz. Around the 1970s non-normal logics were associated with the problem of omniscience in the epistemic interpretation of modal logic, especially in the work of Jaakko Hintikka. The point is that in epistemic logic one would like to avoid situations such as if an epistemic agent believes in the truth of a set of formulae then he also believes in the truth of every logical consequence of this set.

Bonanno's choice is to transfer the non-normality to the information operator. The effect is that if an agent is informed of a formula he is not necessarily also informed of the logical consequences of this formula. However; to be informed of the truth of a proposition means in this approach to have been able to consider all the cases of the truth of this proposition.

The problem with this strategy is that the belief operators continue to assume omniscience. My proposal is to conceive the belief operators as non-normal too, but in some kind of what has been called *conditional necessity*. Elsewhere I developed a non normal dialogic motivated by a kind of abduction which seems to be the counterpart of the project called inverse mathematics. In this interpretation of non-normal modal logic the fact that the law of necessitation does not hold is understood as implementing the

idea that no logically valid argument could be proven in such systems to be unconditionally necessary (or true in any context and logic).

Anyway if we have a set of contexts, how are we to recognise those underlying a standard logic? The answer is clear in modal dialogics if we assume that the players can not only choose contexts but also the (non-modal) logic which is assumed to underlie the chosen context. In this interpretation the Proponent fixes the standards, i.e. determines which is the (non-modal) standard logic underlying the modalities of a given context. However under given circumstances the Opponent might choose a context where he assumes that a (non-modal) logic different from the standard one is at work. Now, there are some natural restrictions on the Opponent choices. Assume that in a given context **O** has *explicitly conceded* that **P** fixes the standards. In other words, the Opponent concedes that the corresponding formulae are assumed to hold under those structural conditions which define the standard logic chosen by the Proponent: we call these contexts *normal*. Thus, **O** has conceded that the context is normal – or rather, that the conditions in the context are normal. In this case **O** cannot choose the logic: it is **P** who decides which logic should be used to evaluate the formulae in question, and as already mentioned, **P** will always choose the logic he has fixed as the standard one. That is what the concession means: **P** has the choice.

Contradictions and non normality

Bonanno's paper deals with new information which contradicts some earlier belief in a typical AGM style: The new information will produce a revision of the old belief. But the normality of the belief operators in combination with the axioms may produce some awkward situations. Recall the minimality axiom: It says that beliefs should be revised in a logically minimal way, in the sense that no new beliefs should be added unless they are implied by the old beliefs and the information received. Assume that at a given state wi the epistemic agent is informed that φ is true though he believed before that φ, from the point of view of the given state, is never the case. Then for any formula φ and the minimality axiom the conditional $\varphi \rightarrow \varphi$ is trivially true at every state which the epistemic agent considered initially to be possible

from the point of view of w_i. Furthermore the assumption of normality excludes cases where we could come to know that the logical consequence of some set of premises contradicts some of our premises.

The mentioned above proposal to redefine the belief operators to handle cases of counterlogicals can avoid such situations. But here too we would like pe to allow some kind of revision rather than only delimitating the conflicting regions as the counterlogical approach does. It is at this point where I think adaptive logic could be of help. Logical revision of logical omniscience problems could then be combined with information revision. But this is only the suggestion of a proposal and I will leave it there.

If we allow the belief operators to be the subject of logical changes too we will still have a controlled and sensible revision strategy, without the inhuman assumption of omniscience. I think this is Peirce's insight regarding the notion of abduction. Abduction is, according to Peirce, about being surprised by new knowledge which can change our views, and the fact that this can happens in science is Peirce's lesson. To be able to change is be able to learn in ever new patterns of reason about our wonderful ever changing world. *Long live Peirce!*

What is abductive method?

Pasi Pohjola
Department of Social Sciences and Philosophy
University of Jyvaskyläa
pasipoh@cc.jyu.fi

1. Peirce and the methods of science and philosophy

In this paper I wish to address the question 'what is abductive method?' in a very general manner. It is a well known fact that C. S. Peirce held abduction as the third form of reasoning, in addition to deductive and inductive reasoning. Peirce also emphasized the relevance of abductive reasoning in processes of scientific discovery. The initial approach to abductive reasoning in this paper is related to a question how Peirce's conception of abduction, his methodological writings of scientific method and nature of philosophical disciplines relate to each other. The intention is not to offer a scholarly investigation of these issues found from Peirce's writings, but to examine whether there can be a general description of abductive method. The argument here is that in order to understand abducive reasoning there are two issues that ought to be taken into account. The first one is fundamental preconditions (of thought) for making propositions about the world, and the second one is the constitutive features of meaning of propositions expressing thoughts.

By the works of N. R. Hanson and more recently by works of L. Magnani and P. Thagard, abductive reasoning has become familiar in the context of scientific discoveries. Also Peirce himself saw that abductive reasoning is a type of reasoning that manifests itself in the process of scientific discoveries. For example, when discussing about Kepler's work, Peirce emphasizes frequently the essential role of abductive reasoning (retroduction) in the course of Kepler's discoveries (CP 1.65 – 74). In the context of

Kepler's discoveries abductive reasoning is demonstrated to be the central method of scientific inquiry. It is also something that is belonging to the heart of Peirce's scientific methodology. In addition to the developments of methodology and classification of sciences, Peirce also had proposals about methodology of logic and philosophy (Haaparanta 2001). It can be said that in Peirce's work *abduction is in part a method of science* and the investigation of abductive reasoning relates to philosophical disciplines (including grammar and logic).

According to Peirce's classification, philosophy divides into three different categories of diciplines. The order of these diciplines is roughly based on the generality of the investigation. In Peirce's classification the first and most fundamental category of philosophical diciplines is *Phenomenology* that investigates the fundamental features of our experience and cognition. It can be said that Peirce's phenomenology investigates most common and general features that constitute our experiences. Second category is *Semeiotics* which can be said to be investigation of how signs and symbols or the conceptual apparthus are related to thoughts and experiences. Third category of philosophical diciplines is *Metaphysics* which is manifestation of these fundamental features of thought, signs and logic investigated in the two previous categories of diciplines.

The most fundamental dicipline of philosophical study is *Phaneroscopy*, the description of the *phaneron*, which belongs to the category of Phenomenology. Peirce writes that "by the *phaneron* I mean the collective total of all that is in any way or in any sense present to the mind, quite regardless of whether it corresponds to any real thing or not." (CP 1.284). According to Pape's interpretation, Peirce's idea in his categorization of philosophical disciplines is that *Phaneroscopy* provides theoretical and methodological principles for *Semeiotics* and both *Phaneroscopy* and *Semeiotics* provide theoretical and methodological principles for *Metaphysics* (Pape 1990).

Maybe the best known issues of Phaneroscopy are Peirce's conceptions of firstness, secondness and thirdness. This trichotomy can also explicate Peirce's *Phaneroscopy* in more detail and show how it relates to abductive reasoning in Peirce's work. Without going into details of this tichotomy,

it can be argued that Peirce's idea of thirdness is a preliminary condition of thought for what is more recently described as the general logical or grammatical form of propositions (f. ex. by G. Frege and L. Wittgenstein). Thirdness is also something that preoccupies our perception and concept formation prior to any reasoning together with firstness and secondness. The link of these issues to abduction is that hypothesis suggestion is a suggestion of a possible proposition with a certain propositional form.

The issues of investigation belonging to disciplines under Peirce's conception of Semeiotics (semiotics) are provided with methodological and theoretical principles from Phaneroscopy. The diciplines belonging under the heading of Semeiotics are, according to Pape, speculative grammar, critic and methodeutic, and they can be described as a formal doctrine of signs (distinguished from concrete investigation of meanings of signs and sentences). Without going into discussions about Peirce's semiotics and the theory of signs, it has to be emphasized how the ideas of phaneroscopy, especially thirdness, are reflected and represented in logic and grammar. Peirce himself is explicit about the role of thirdness (including firstness and secondness) in having meaningful thoughts and constructing propositions. According to Peirce, thirdness is a general feature of conciousness that represents itself in logical or grammatical forms of propositions.

Following these general ideas of two Peircean topics, Phaneroscopy and formal doctrine of signs, the question concerning whether there is an abductive method is taken into discussion. N. R. Hanson has stated that abduction relates to the conceptual aspect of problem solving, and this idea is discussed in light of Perice's ideas about philosophical diciplines, especially how they relate to the construction of meaning. After laying forehand some conceptual tools from Peirce's philosophy the intention is to describe certain preliminary and general conditions of abductive reasoning that have essential role in constructing thoughts and in conceptualizing observable phenomena. Finally it is argued that these features of abductive reasoning discussed in the paper present preliminary conceptual conditions for reasoning processes, and that these features describe foundations of a method for suggesting novel hypotheses.

2. Phaneroscopy and forms of description

In Peirce's system of philosophical disciplines phaneroscopy is a study of most fundamental features of thought. In his discussion of this most fundamental discipline of philosophical study, he writes that:

> "I invite you to consider, not everything in the phaneron, but only its indecomposable elements that is, those that are logically indecomposable or indecomposable to direct inspection." (CP 1.288)

According to Peirce, phaneroscopy brings about the most fundamental categories of thought. As it is stated above, these categories are a trichotomy, typical to his philosophy, and he calls these categories as firstness, secondness and thirdness. In his Harvard lectures on Pragmatism he claims that firstness refers to *Quality of Feeling*, secondness to *Reaction* and thirdness to *Representation*. In his previous Havard lecture dealing with phenomenology he claims that a proper theory of categories is such that it has its representative in the discipline that has to do with signs and logic.

In these seven Harvard lectures on Pragmatism Peirce also considers the question of meaning and what ought to be a theory of meaning. Where Phaneroscopy is about general categories of thought and logic is about proper reasoning and forms of representations, meaning (of signs) deals with the content of representations expressed in signs and in certain forms defined by logic and grammar. According to Peirce, the very possibility of meaning, especially the meaning of propositions, is dependent on certain general categories of thought (thirdness) and its representative logical or grammatical forms. This similar kind of idea of meaning is existent also in Edmund Husserl's early work and in Ludwig Wittgenstein's work. (Peirce 1998)

In his early work, Husserl speaks of pure grammar of meaning as a theory of meaning forms. These meaning forms lay an ideal framework which actual language fills up (Husserl 1970, 518-526). Similar kind of an idea can be found from Wittgenstein's conception of grammar, especially when related to his remarks on hypotheses (Pohjola 2006). Wittgenstein claims that grammatical investigation "may be called an analysis of our forms of expres-

sion" and this investigation "is not directed to phenomena, but, as one might say, towards the '*possibilities'* of phenomena" (1995, § 90). These forms as the possibilities of phenomena, or meaningful propositions, give the possible framework for making meaningful linguistic expressions about phenomena.

It is suggested here that what Peirce introduces under the name of Phaneroscopy, has resemblance at least in its basic features, with what Wittgenstein thought philosophical investigation to be, especially in his latter work. Peirce writes that:

> "English philosophers have quite commonly used the word *idea* in a sense approaching to that which I give to *phaneron*. But in various ways they have restricted the meaning of it too much to cover my conception […], besides giving a psychological connotation to their word which I am careful to exclude." (Peirce CP 1.285)

This reference to the word *idea* and the also to the restrictedness of the meaning of it has similarities with Wittgenstein's idea in his philosophical investigation. Wittgenstein states that philosophy should not be targeted to the phenomena itself (restricted ideas) but to the possibilities of phenomena. It can be said that in order to do philosophy in Wittgenstein's sense or to practice Phaneroscopy, one has to have something as begin of investigation that Peirce in the quotation described as ideas. The difference between investigations targeted on ideas and Phaneroscopy is that the latter starts from the ideas i.e. form what is presented to our mind, or as Wittgenstein would say, of thoughts and understanding.

As it has been already argued previously (see also Pohjola 2006), Wittgenstein's philosophical method is targeted to the forms of description. Also Peirce's idea of *phaneron* includes what can be called as forms of description which is also the very target of Wittgenstein's philosophy. What Peirce says of his Phaneroscopy is that it is a study of very general features of the phaneron and the description of these features relies on each general category that is found. Also he states that "[i]t does not undertake, but sedulously avoids, hypothetical explanation" (Peirce CP 1.285-86). In similar way Wittgenstein says that philosophy does not make any hypotheses, it only de-

scribes (Wittgenstein 1995, § 109). Peirce also comes to emphasize an aspect of foundations, also found from Wittgenstein's philosophy, which can be labelled as the *phenomenological* aspect, by stating that Phaneroscopy "simply scrutinizes the direct appearances, and endeavors to combine minute accuracy with the broadest appearances" (Peirce CP 1.287).

Herbert Spiegelberg, in his famous paper 'The Puzzle of Wittgenstein's Phänomenologie (1929-?)', discusses Wittgenstein's use of the concept phenomenology in his manuscripts and also some ideas existent in his work that relate to phenomenology (see Spiegelberg 1982, 202 – 217). According to him, Wittgenstein distinguishes between two different, though related conceptions of phenomenology in his writings. On one hand, Wittgenstein refers usually with his use of adjective *phänomenologisch* to descriptions of phenomena given in experience. The essential tool for such a priori phenomenology is phenomenological language, of which Wittgenstein seems to speak of also in two different ways. H. Spiegelberg refers to these two different conceptions of phenomenological language with terms primary and secondary phenomenological language. Primary phenomenological language is the language of *Tractatus*, declared later dispensable by Wittgenstein. Secondary phenomenological language, on the other hand, is "based on the essentials of our language […] which allows us to grasp the "essence" of the described" (Spiegelberg 1982, 205).

The other conception of phenomenology that Wittgenstein seem to have treats phenomenology as something dealing with possibility of phenomena, not with actual phenomena. The phenomenology is directed to the language of description, used for giving descriptions of our experiences, instead of being directed to describing the given. Spiegelberg says that "[i]t's function is to give framework which can be used in the description of concrete facts" (Spiegelberg 1982, 205). It examines, stated in Wittgenstein's terms, the grammar of our descriptions. Combined together with the idea of secondary phenomenological language, language of phenomenology is a language (or a language–game), which we use for describing our grammar of description (ways of possible descriptions of phenomena). This second conception and secondary language can be seen in a sense as a meta-language,

or as a language–game used for describing language. Actually Spiegelberg concludes that "[i]n this sense it would be a meta-phenomenology rather then a phenomenology" (Spiegelberg 1982, 205). This other conception of phenomenology, a conception that strikes Spiegelberg as completely original, can be seen as a part of grammar and grammatical investigation, central theme in Wittgenstein's latter philosophy. (Spiegelberg 1982, 205)

This second conception of phenomenology combined with the secondary phenomenological language, and the relation of these concepts to Wittgenstein's conception of grammar, are a good motivation for investigating phenomenological aspect of Wittgenstein's latter philosophy, at least in means of revealing something relevant of his philosophical method. It's been argued f. ex. by M. B. Hintikka, J. Hintikka and H. Spiegelberg that a kind of idea of phenomenology (or meta-phenomenology) remains in Wittgenstein's thinking, although he abandons the first conception and the possibility for primary phenomenological language (Hintikka. & Hintikka 1986, 137 – 175; Hintikka 1996, 55 – 77 & 191 –240; Spiegelberg 1982, 202 – 217). This is also the general similarity that Wittgenstein has to C. S. Perice's idea of Phaneroscopy.

Although there seems to be relevant reasons to argue for the survival of Wittgenstein's second conception of phenomenology in his latter writings, a part of the puzzle still remains. Where and why does the terms Phänomenologie and phänomenologisch vanish from his writings? One suggestion that can be made, following ideas of J. Hintikka and H. Spiegelberg, is that his use of Phänomenolgie in the second conception sense is in a way metaphorical, in relation to his first and strict sense. In this way of arguing, it can be said that Wittgenstein is referring literally with the terms Phänomenologie and phänomenoloisch to something which he himself declared dispensable. The metaphorical use, on the other hand, plays a demonstrative role in his change of views and the new conception is defined as grammar (examples of this can be found f. ex. from a manuscript title Phänomenologie ist Grammatik) (Wittgestein 1975b, § 1).

Although the term Phänomenologie vanishes, for a reason or another, from Wittgenstein's terminology, the interest for investigating phenomeno-

logical issues as an access to his philosophical method in his latter work remain. Even for a phenomenologist. H. Spiegelberg has told us that a German phenomenologist Alexander Pfänder has thought "that phenomenology proper presupposes a conceptual analysis under the name of "clarification of meaning (Sinnklärung)" " (Spiegelberg 1982, 83 – 84). This Pfänder's idea is quite close to the idea of Wittgenstein's 'vanished' and very probably remained idea of phenomenology as a philosophical method. Wittgenstein states that the central method of philosophy is linguistic analysis or the analysis of language-use. The idea of philosophy is to clear away problematic uses of language by description. Although Wittgenstein discusses of description of language-use, one should keep in mind the fact that this investigation is not targeted to phenomena (mere actual language use) but to the possibilities of phenomena, and that the investigation can be said to be a grammatical one (Wittgenstein 1995, § 90).

3. Some problems of abductive inferences

In discussions concerning inference to the best explanation (IBE), Peter Lipton has engaged into a debate with Suppe and van Fraassen about the very possibility of identifying some inference as an inference to the best explanation (see Lipton 1998). In short, the main criticism against IBE that van Frasasen (1989) has made is that in science one does not typically choose to infer the best explanation, but only some explanation. In the case of discoveries or more accurately in suggesting explanatory hypotheses, van Fraassen's claim is worth noticing. If we are suggesting a hypothesis that could explain anomalious phenomena, we cannot know which hypothesis is the best one. At least we cannot have any epistemic reasons for granting some hypothesis as the best one, because this would presuppose knowledge of what is the case we are trying to explain. But if we already know what is the case, then there isn't any need for suggesting explanatory hypothesis.

The critisicism of IBE from the Semantic approach by van Fraassen is epistemological and by recongizing these epistemological pitfalls of IBE Peter Lipton has suggested that an inference to the best explanation should be thought of as the 'loveliest' explanation. In general, this move draws

the source of evaluative criteria away form epistemic tribunal to something else. This is also the difference between the standpoints of the authors of this quarrel. Where Suppe and van Frassen argue against IBE by critisizing the possibility of empirical validation and evaluation of these inferences, Lipton's argument for IBE is ontological by stating that the criteria for the best explanation is something else then empirically recongizible likeliest explanation.

Although Lipton's argument for IBE can be problematic for his account, there in general is a certain overall idea that is adopted here from his account. Because of the problems that van Frassen has shown to exist in IBE, epistemic criteria, at least alone, cannot be the source for suggesting some hypothesis rather then other. It is suggested here that there are certain criteria that can be drawn from previous discussions in this paper that can be thought of as a prerequisite for any epistemic criteria. These are, it is here claimed, features of our thought that Peirce is discussing about in his phaneroscopy and something that Wittgenstein is discussing about when he talks of possibilities of phenomena. The common feature of these discussions is that these discussions are about prerequisites for meaning and meaningful propositions, such as explanatory hypoteses. In what follows, some of these features are demonstrated by a common sense example, using Sherlock Holmes as the demonstrative character.

In a first chapter of his novel 'Sign of Four', Arthur Conan Doyle gives a description of a deductive method used by Sherlock Holmes in his investigations. In the story, Sherlock Holmes discusses of his method of investigation with Dr. Watson, calling it pure logic and deduction. Holmes comes to discuss his method after he has made an investigation about the origin of Dr. Watson's watch. When Watson accuses Holmes of making a pure guess, Holmes concludes that "[n]o, no: I never guess. It is a shocking habit–destructive to the logical faculty. What seems strange to you is only so because you do not follow my train of thought or observes the small facts upon which large inferences may depend "(Doyle 2001, 40). Two things in this story can be found to be of the most interest for an investigator of abductive method: The role of observations and the ability to extract something

relevant to the inferences from observations and of course the inference it self, because Holmes denies that he is guessing.

Abductive reasoning is said to appear constantly in our daily practices, such as in problem solving. As it is remarked previously, one of the best known figures in detective stories who use abduction intensively is Sherlock Holmes (Eco & Sebeok 1988). His deductions in his adventures reveal to be abductive reasoning. For example, when solving the case of the speckled band, Holmes obviously doesn't have all necessary facts (premises) for deducing that the speckled band is a poisonous snake etc. Eventually he comes to an obvious or best possible conclusion, and he does this not by guessing but drawing something relevant out of his observations. Although he had previously in the same story reasoned wrongly from insufficient data, his data for the conclusive correct inference is not sufficient either, at least for making a pure deduction. (Doyle 2001, 111 – 117) Philip Clayton has characterized this kind of inference as inference to the best explanation (Clayton 1997), which is not deduction but still it has empirical validly, i.e. it is correct inference of the case.

It can be also argued that Sherlock Holmes is, without acknowledging it, applying certain kind of phenomenological method or conceptual analysis in his inquiry and as a part of reasoning. When he reminds Watson not to look at the obvious, but behind the obvious, he seems to suggest something phenomenologists call bracketing the natural attitude towards the observable. In the analysis of observable in the abductive reasoning, it seems that Holmes draws his attention to meaning and the meaning constituting contribution. The solution for cracking the case has to be meaningful, although not necessarily obvious and Holmes seems to consider what solutions could make sense, although he cannot right away know if they really do. As in every creative process, such as discoveries or solving a case, the conclusion has to be meaningful to be successful, it cannot be nonsense. Therefore Holmes is inclined to penetrate to the possible phenomena searching alternatives that could explain the case and this has, in the view taken here, evident similarities to what is in question in discoveries and abductive reasoning in general.

One necessary step in solving a mystery can be located in the fact that in some phase of his investigations Holmes comes to an understanding of what is the case in whole. This has similarities to what Husserl describes as eidetic reduction and *Wesensshau*, success of disclosing the invariant features of the case. But the problem for a philosopher in the analysis made by Holmes so far lies in the fact that this disclosure doesn't reveal (all the) relevant understanding of the possible hypothesis or the solution of the case. This analysis has only come backwards from observations to very edge of Metaphysics in Perice's categories of philosophical disciplines. This is one reason why Husserl in his philosophy turned to the transcendental and transcendental reduction, as Zahavi argues (Zahavi 2003, 52-53).

The need to go further in the analysis is an idea that also comes close to what Peirce says about Phaneroscopy. One should not consider everything in the phaneron but only most invariant or indecomposable features. It is a step from understanding what is common features *in* our perception to what is common *to* (or constitutive of) our perceptual experience. Also Wittgenstein's claim that hypothesis generation is not about actual phenomena but about possibilities of phenomena. This characterization fits also to Holmes' inquiries, especially to the phases where Holmes is short of evidence and he has to consider what could be the possible solution of the case (see more detailed discussion below).

The reason for discussing these issues is that in abductive reasoning, there are certain non-epistemic features that have a central role in evaluation of possible hypotheses. Because the case in abductive reasoning is not about logically sound deductive inferences, there needs to be some other kind of criteria than logical criteria for the success or correctness of the inference. One possible suggestion made in discussions of inference to best explanation is that the inference is empirically correct, which is a criteria applied for conceptual inferences and inferences in epistemic logics. But there are certain limitations for applying this kind of epistemic criteria to inferences. When discussing about conceptual inferences, such as these inferences here, the epistemic criteria is dependent of some meaningfulness criteria and epistemic criteria can be applied only after inference is made (post inference

criteria). The conclusions of the inferences ought to be meaningful before any empirical criteria can be given. Therefore it is argued that a kind of phenomenological investigation or analysis directed to phenomena about this subject, which is targeted on the meaning constitution, is relevant.

4. Grammar, hypotheses and possibilities of phenomena

Wittgenstein's central claim in *Tractatus* was that the world is the totality of facts. This is also his central ontological conviction and it can be argued that it is something existent in his latter philosophy. The changes in his philosophy were changing immediate object – name relation to the idea of language-game and extending the conception of grammar from logical grammar to the grammar of every day language. Although certain authors see Wittgenstein's change of mind as a fundamental turn from the language of *Tractatus* to the language of *Philosophical Investigations*, this is obviously not all that is the case. It is held here as a fact that Wittgenstein shares similar kinds of conceptions both in *Tractatus* and in *Philosophical Investigations* that are very central to his work.

In his latter work, especially in *Philosophical Investigations*, Wittgenstein abandons the immediate name-object relation of *Tractatus*. Some commentators have also seen this as evidence for the fact that Wittgenstein is abandoning the idea of logical form of propositions. Although Wittgenstein abandoned the immediate object-name relation and replaced it with the conception of language-game, it can be argued that his idea of logic has survived and is a part of his latter conception of grammar, as the passages quoted above seem to suggest (Bloor 1983, 22; Hintikka & Hintikka 1986, 190 – 199). What Wittgenstein abandons from his earlier work is the idea of immediate name-object correspondence. The idea of logic itself survives and becomes developed as a part of what is known as grammar in his latter philosophy. The remarkable change in his thoughts about meaning is a change from immediate object-name relation to language-games that are 'played' by the rules of grammar. In his latter work Wittgenstein argues that using language is playing different language-games and the meanings of expressions are dependent on the contexts they are used in.

The logical constraint of reason, in the sense that our experiences are meaningfully structured and that we can have propositions about the world, has produced ideas that logic is somehow sublime in our thinking (McMullen 1989). In §§ 89 – 90 of *Philosophical Investigations* Wittgenstein discusses the issue of how logic is thought to be somehow sublime.

> "In what sense is logic something sublime? For there seemed to pertain to logic a peculiar depth – a universal significance. […] We feel as if we had to *penetrate* phenomena: our investigation, however, is directed not towards phenomena, but, as one might say, towards the '*possibilities*' of phenomena. We remind ourselves, that is to say, of the *kind of statement* we make about the phenomena." (Wittgenstein 1995, §§ 89 – 90)

In these passages Wittgenstein comes to criticize the idea that logic itself is something sublime. These passages in *Philosophical Investigations* are also a part of his criticism of *Tractatus*. The criticism against *Tractatus* is not targeted at logic as such, but against the idea that there is a logical primary language in contrast to secondary everyday language. The interesting remark for ontology in this general critical comment on the sublime logic view is that Wittgenstein is reminding that the investigation should be targeted at the ways we make plausible statements about phenomena, not necessarily something sublime. In the same passages Wittgenstein writes that "the essence of our investigation [is] that we do not seek to learn anything *new* by it. We want to *understand* something that is already in plain view". (Wittgenstein 1995, § 90) So the investigation he is suggesting ought to be targeted at the logic of our description, as it is apparent in the ways we make descriptions, not at something sublime. It should be targeted at the ways we construct logically structured plausible propositions of the world.

As it was said above, the conception of grammar replaces the more strict conception of logic as the structure or form of our thoughts. Maybe it would be actually more accurate to say that grammar has replaced logic as the preconditions of meaningful thoughts. In *Philosophical Investigations* Wittgenstein actually writes that "[e]*ssence* is expressed by grammar" and that "[g]rammar tells what kind of object anything is" (Wittgenstein 1995, §§ 371, 373). It is, then, grammar that lays down the rules of description, i.e. the preconditions

for making possible descriptions and thus having thoughts. (Pohjola 2006) In here also a similar kind of way of thinking between *Tractatus* and *Philosophical Investigations* can be noticed. Only the definite plea for logic has been replaced with the notion of grammar.

Wittgenstein does not refer with the conception of grammar merely to something that we are acquainted with, i.e. a set of explicit rules for sentence structure etc. In stead his conception of grammar is essentially connected to meaning and to the possibility of meaningful expressions. It, of course, includes the syntactical rules of sentence formations but it is also connected to the context of use. The grammar of some expression is not just about what kind of a sentence is usable but also what kind of combination of words is applicable. For example, a statement 'the ball is a square' would not be a possible and thus correct application of words in a sentence used in the context of girls and boys playing football in a park. We can, thus, imagine a situation where this would be applicable. In an adventure of Donald Duck created by Carl Barks in 1949 called 'Lost in the Andes', Donald and his nephews discover a lost culture in the Andes where chickens lay square eggs and even the local boys kick a ball that is square.

The core idea of linguistic analysis as a part of Wittgenstein's philosophical method can be demonstrated through some initial ideas of a theory of meaning proposed by J. J. Katz. This can also demonstrate the necessary relation between structure and context. Jerrold Katz has argued, very much against the 'meaning-is-use' view of meaning, that Fregean semantics is an essential part of the theory of meaning. According to Katz, the theoretical investigation of meanings of linguistic expressions should take into account two perspectives: *Fregean semantics* and *contextual pragmatics*. Fregean semantics is concerned with the ways how meaning is dependent on the logical and grammatical structure of sentences. (Katz 1981) The fundamental idea of Fregean semantics is described by Peter Hacker by a slogan "words have meaning only in a context of a sentence" and the central issue is the investigation of the relations of formal structures and meaning (Hacker 1979). On the other hand, contextual pragmatics in the theory of meaning is a shift of investigation from abstract sentence meanings to contextually determined

meanings of language use. Contextual pragmatics investigates how well structured sentences are used in different contexts. For Katz, both of these are essential features of the theory of meaning. Katz defines pragmatic theory, including Fregean semantics, in a following manner:

(P) $\quad PRAG\ (D(S_i), I(C(t))\ ) = \{R_1, \ldots, R_n\}$

The arguments in (P) are: $D(S_i)$, which is the full grammatical description of a sentence S_i and $I(C(t))$ is a specification of all the relevant information about the context C, where certain token t of S_i occurs . The output of PRAG is a set of sentence readings $\{R_1, \ldots, R_n\}$. (Katz 1977, 16)

Although Katz's theory has been discussed critically in many respects (Bach & Harnish 1982; Jackendoff 1981), there is a moral found from Katz's story. When we make statements about the world, i.e. when we give descriptions of the phenomena, the statements are always made in a context and also these statements have certain structure. It can be even argued that the kinds of possible statements are bounded to only some possible forms and also the kinds of possible statements are bounded by the existing context. Therefore both aspect of meaning are essential in the investigation. According to Katz the meaningfulness criteria for meaningful expressions of thougts ought to be considered from two interrelating perspectives: From the perspective of Fregean semantics and from the perspective of context constituting the meaning.

In 1930, in a discussion with Waismann and Schlick, Wittgenstein makes a claim about the nature of hypotheses. He distinguishes between *propositions* and *statements* on the one hand, and *hypotheses* on the other. The statements and propositions are descriptions of states of affairs that can be true or false, which is of course very common way of thinking about statements or propositions. On the other hand, hypotheses are not descriptions in the same sense that propositions are. According to Waismann, Wittgenstein has said that a "hypothesis is not a statement, but a law for constructing statements." (Waismann 1979, 99) Later in his manuscripts there is a passage where he claims that a hypothesis is a logical structure, to which the rules of representation hold (Wittgenstein 1975b, 283). According to Noë, the

distinction Wittgenstein makes between hypotheses and statements is that hypotheses and statements have a different kind of formal (or grammatical) relation to reality (Noë 1994, 14 & 34).

What Wittgenstein is saying about hypotheses is that they are not descriptions of phenomena in the same sense as propositions or statements. Wittgenstein thinks that hypotheses are descriptions of possibilities of phenomena that can become propositions or statements. This idea of possibility of phenomena relates to what Wittgenstein describes as grammar and grammatical investigation (Wittgenstein 1995, § 90). Grammatical investigation is not, in the context of hypotheses, targeted to the phenomena itself, but at the forms of expression that are used for describing the phenomena. Hypothesis suggestion is about the possible ways of conceptualizing the phenomena, ways of making description. One thing that here should be noted is that what Wittgenstein means with his terms grammar and grammatical is not just the syntactical rules of language structure, but also the rules of meaningful use of language (Fregean semantics and the context of language use).

What the difference in descriptions of phenomena and descriptions of possible phenomena is can be demonstrated by referring to the stories of Sherlock Holmes once more. The typical quarrel between Holmes and Watson is about what to 'see' in the cases. When Watson describes what is obvious and makes hasty conclusions about the case, Holmes suggests his friend to see beyond the obvious. It can be said that Watson is describing the phenomena and relying on his everyday conception of things. Holmes, on the other hand, is going beyond the obvious and seeking for possible explanations for the case. It can be said that Holmes' method of inquiry targets to the logical, conceptual and epistemic possibilities that can be the case.

5. Conclusions: Does abductive method really exist?

In previous sections certain fundamental philosophical issues relating to the nature of hypotheses and hypothesis generation are discussed. The intention of these discussions has been to show what kind of questions and issues relate to abductive method. The question whether there is an abductive

method can be answered only after answers to questions concerning criteria of inferences is answered. Above it was argued that epistemic criteria alone, at least in the sense that it needs empirical validation, cannot function as criteria for choosing and suggesting hypotheses. In addition to epistemic criteria some additional criteria were suggested. These criteria can be labelled as phenomenological or transcendental in Kantian sense, but following Peirce these criteria are strongly connected to signs or to the conceptual aspect of consciousness.

Relying strongly on Wittgenstein's philosophy the nature of hypotheses was discussed. According to Wittgenstein characteristic of hypotheses is that they are about possibilities of phenomena and not actual phenomena. Suggesting hypotheses relies on certain rules of description that distinguishes the possible forms of suggestions from improper ones. This idea of possible descriptions relates strongly to Wittgenstein's conception of grammar in his *Philosophical Investigations*. This conception of grammar, in addition to rules of possible logical or syntactical forms of description, includes also the pragmatic aspect of rules. In certain situations and in certain contexts, only certain kinds of descriptions are possible. It can be said that this conception of grammar relates to both form and content of propositions.

Turning now to the initial question of this paper already mentioned in its title about the existence of abductive method, one can say that Peirce himself believed in it, although he wasn't very explicit about it. One certain characteristic of abductive method is that certain kind of reduction or change of perspective that directs the investigation to more general features of phenomena is taken. Also suggestion of explanatory hypothesis is bound by meaningfulness criteria, i.e. the explicated proposition has to have the logical and syntactical form of a proposition and that it has to be understandable. This meaningfulness criterion can be investigated by linguistic analysis, but as it was argued through references to authors such as Peirce, Wittgenstein and Husserl, the search for possible explanatory hypotheses goes beyond meanings of existing conceptualizations and related epistemic criteria.

To conclude, the question addressed in the beginning of this paper can be given, al least in my view, an initial positive answer. Reasoning that bring

about novel conceptualizations of phenomena (which can be identified as abductive reasoning) has certain general features and characteristics that can be seen as reasons for calling it an abductive method. The down side is that these features and characterizations are very general and therefore they cannot provide means for specifying the method to more applicable form. Neither do they provide means for composing a guide book of abductive method for scientist. So, although the question was given a positive answer, the results of analysis are philosophical and still far away from everyday practices of scientist.

References

Bach, K. & Harnish, R., 1982, Katz as Katz can, *The Journal of Philosophy* 79: 168-171.

Bloor, D., 1983, *Wittgenstein. A Social Theory of Knowledge*, Macmillan Education, London.

Clayton, P., 1997, Inference to the best explanation, *Zygon* 32:3, 377-391.

Doyle, A. C. 2001, *The Complete Sherlock Holmes*, Geddes & Grosset, New Lanark.

Eco, U. & Sebeok, T. A. (eds.) 1988, *The Sign of Three. Dupin, Holmes, Peirce.* Indiana University Press, Bloomington.

Haaparanta, L. 2001, On Peirce's methodology of logic and philosophy, in M. Kiikeri & P. Ylikoski (eds.) *Explanatory Connections. Electronic essays dedicated to Matti Sintonen*, http://www.valt.helsinki.fi/kfil/matti/

Hacker, P. M. S., 1979, Semantic holism: Frege and Wittgenstein, in: *Wittgenstein, Sources and Perspectives*, C. G. Luckhardt, ed, Harvester Press, Sussex, 213-242.

Hintikka, J., 1996, *Ludwig Wittgenstein. Half-Truths and One-and-a-Half-Truths. Selected Papers 1*, Dordrecht, Kluewer.

Hintikka, M. B. & Hintikka, J., 1986, *Investigating Wittgenstein*, Basil Blackwell, Oxford.

Husserl, E., 1970, *Logical Investigations Vol 2.*, J. L. Findlay, ed, Routledge & Kegan Paul, London.

Jackendoff, R., 1981, On Katz's autonomous semantics, *Language* 57(2): 425-435.

Katz, J. J., 1977, *Propositional Structure and Illocutionary Force*, Harvester Press, Sussex.

Katz, J. J., 1981, Literal meaning and logical theory, *The Journal of Philosophy* 78: 203-233.

Lipton, P., 1998, The best explanation of a scientific paper, *Philosophy of Science* 65, 406 – 410.

McMullen, C., 1989, In what sense is logic something sublime?, *Noüs* 23: 35-61.

Noë, R. A., 1994, Wittgenstein, Phenomenology and what it makes sense to say, *Philosophy and Phenomenological Research* 54: 1-42.

Pape, H. 1990 Charles S. Peirce on objects of thought and representation, *Nöus* 24:3, 375-395.

Peirce, C. S. 1960 *Collected Papers Vol. I-2.*, C. Hartshorne and P. Weiss (eds.), Harvard University Press, Cambridge.

Peirce, C. S. 1998, *The Essential Peirce. Selected Philsophical Writings 1893-1913*, Indiana University Press, Bloomington.

Pohjola, P., 2006, Abductive reasoning and linguistic meaning, *Logical Journal of IGPL* 14:2, 321 – 332.

Spiegelberg, H., 1982, *The Context of the Phenomenological Movement*, Martinus Nijhoff, London.

van Fraassen, B., 1989, *Laws and Symmetry*, Oxford University Press, Oxford.

Waismann, F., 1979, *Wittgenstein and The Vienna Circle*, Basil Blackwell, Oxford.

Wittgenstein, L., 1961, *Tractatus Logico-Philosophicus*, Routledge & Kegan Paul, London.

Wittgenstein, L., 1975, *Philosophical Grammar*, Basil Blackwell, Oxford.

Wittgenstein, L., 1995, *Philosophical Investigations*, Blackwell, Oxford.

Zahavi, D., 2003, *Husserl's Phenomenology*, Stanford University Press, Stanford.

Abductive Logic of Discovery with Distributed Means

Sami Paavola
Centre for Research on Networked Learning and Knowledge Building
University of Helsinki
sami.paavola@helsinki.fi

Introduction

In the 1950s and 1960s, N. R. Hanson (1958, 1961) suggested that Charles S. Peirce's abduction may serve as a basis for a "logic of discovery". Hanson's aim was not any formalistic system of logic, yet he maintained that Peirce's basic formulation of abduction provides a conceptual apparatus for understanding those processes through which new ideas and theories are discovered, and new explanations of phenomena are searched for (Hanson 1958, 2-3). Most philosophers at that time maintained that such conceptualization is *not* possible. According to them processes of discovery can not be analyzed conceptually and with philosophical means, only those ways by which the idea is justified. In many ways this is still the prevalent conception.

Peirce's own writings concerning abduction have left room for various interpretations, which bear upon its possible use in accounting for discovery. Abduction as a third main mode of reasoning seems to imply that Peirce was aiming at finding a logic of discovery. On the other hand, in his later writings, Peirce emphasizes that abduction is very close, or even the same as a guessing *instinct* for finding fertile ideas. This seems to imply that there is no *logic* of discovery, but rather discovery is about such things as instinct, intuition, and guessing. Peirce himself, in his later writings, did not make any difference between these two ways of interpreting abduction but subsequent

interpretations have often emphasized either inferential or instinctual aspects of abduction.

It can, however, be argued that these two basic ways of interpreting abduction are not enough if the aim, like Hanson's, is to understand and conceptualize processes of discovery with abduction. They easily depict inquiry in too rationalistic, mentalistic and individualistic terms, as if discovery were something which human beings do just by making inferences within their own mind. In any case, modern ideas concerning *distributed cognition* have forcefully challenged this kind of a picture about human rationality (Salomon 1993; Hutchins 1995). But what happens to abduction and to the idea of logic of discovery if distributed cognition is taken seriously?

In this paper I first briefly delineate Peirce's own development concerning abduction. Then I present what I think was the main idea behind the Hansonian "research programme" on a "logic of discovery". After that, I analyze how the ideas of distributed cognition have been and could be taken into account within abduction. Distributed cognition gives new means for understanding processes of discovery with a broader and more realistic framework than before. In many respects Peirce's ideas fit well with modern ideas about distributed cognition, but these kinds of ideas are *not* very prominent in his specific theory concerning *abduction*. I maintain that distributed cognition is one central direction where abductivist methodology should be developed if the aim is to develop conceptual means for understanding processes where new ideas and hypotheses are generated and developed.

Basic interpretations of Peirce's abduction

Peirce's writings concerning abduction are usually separated into two main periods although the change was not an abrupt one (Burks 1946; Fann 1970; Anderson 1987). The first period can be called a "*syllogistic*" (or "evidential") interpretation of abduction (from 1860s to about 1890s) when Peirce interpreted abduction through various forms of syllogistic reasoning, and as a weak form of an evidencing process. Abduction is then seen as an inversion of Barbara syllogism so that the minor premise is abduced from the conclusion and the major premise. In his later writings (about 1890s on-

wards) Peirce did not abandon this interpretation as such but he emphasized that abduction is an essential element of the first phase of inquiry where ideas are originated for subsequent testing. This later, "*methodological*" approach had a broader view on inquiry than the earlier approach.

The change from syllogistic interpretation to methodological interpretation in itself was not a very large one because in his later writings Peirce still maintained that abduction is a form of reasoning with a similar syllogistic basic structure that he had presented earlier. A clearer change was that in his later writings Peirce maintained that abduction has its basis in an "*instinct*" for finding or guessing fertile hypotheses (CP 7.219, 1901[1]). Peirce presented various rationales for this instinct in his writings during the years (Shanahan 1986; Paavola 2005), but the basic idea was that the success of human beings of finding good hypotheses cannot be explained except by assuming that there is some kind of a tendency of human beings for finding promising ideas. Peirce also proposed that abduction is near to perceptual judgments, so it is akin to perceptual insights (CP 5.181, 1903; Hanson 1958). This has led to difficulties of interpreting how abduction can be both a form of an inference and a sort of an insight (or an instinct) at the same time (Frankfurt 1958, 594; Fann 1970, 35; Anderson 1987, 32,35).

Subsequent interpretations of Peirce's abduction has often emphasized abduction either as an inference or as some sort of an instinct (or an insight) by either treating abduction as a "pure" form of inference without giving any special role for an instinct, or, by interpreting abduction as a sort of an "intuition" for noticing clues and signs and finding solutions for difficult puzzles (see e.g., Eco & Sebeok 1983). Often these interpretations are mixed together without making any clear analysis how they are related. For example, when Hanson developed his notions concerning abduction as a "logic of discovery", he appealed to the inferential structure of abduction, and the way how scientists reason from surprising phenomena to new explanations (Hanson 1961). He did not appeal to any instinct here. On the other hand,

1 'CP v.ppp, xxxx' refers to Peirce's Collected Papers (see Peirce 1931-1958) in a customary way by indicating after CP a volume number (v), a paragraph number (ppp) and the year when the manuscript was written (xxxx).

Hanson connected abduction to "Gestalt" switches, and to "seeing" new kinds of patterns in phenomena (Hanson 1958). Peirce already had maintained that abduction

> "shades into perceptual judgment without any sharp line of demarcation between them; or, in other words, our first premisses, the perceptual judgments, are to be regarded as an extreme case of abductive inferences, from which they differ in being absolutely beyond criticism" (CP 5.181, 1903).

But this does not in itself solve the dilemma when it comes to the logic of discovery. Is abduction supposed to be a form of inference from data to hypotheses, or is it more like a sudden flash of insight based on a mysterious capacity of human beings of interpreting things? If it is both, how are we supposed to connect these two?

Peirce himself never called abduction as a 'logic of discovery'. Still, this kind of an interpretation is easy to make for many formulations which Peirce proposed concerning abduction, especially in his later writings. In his early writings, abduction concerned more justification, and in the later writings discovery (Thagard 1981) although the change is not an abrupt one. In his later period, Peirce maintained that abduction "furnishes all our ideas concerning real things, beyond what are given in perception" (CP 8.209, c. 1905). Peirce also wrote that "[a]ll the ideas of science come to it by the way of Abduction. Abduction consists in studying facts and devising a theory to explain them." (CP 5.145, 1903). Peirce often seemed to emphasize that the instinctual tendency to guess right is the key for explaining how new ideas are generated[2] although there was no sharp contrast with "reasoned" and "instinctive" considerations (CP 7.220, 1901).

My own interpretation is that abductive inference and abductive "instinct" should be separated more clearly than Peirce did in his later writings (Paavola 2005), but both of them should be developed further if abduction is to be understood as a way of conceptualizing actual processes of discovery. But besides these, a methodological treatment of abduction can

2 "… every plank of its [science's] advance is first laid by Retroduction [i.e. abduction] alone, that is to say, by the spontaneous conjectures of instinctive reason". (CP 6.475, 1908)

be expanded to those areas which are emphasized within *distributed cognition*. Although Peirce's and Hanson's basic models of abduction were different from distributed cognition, their approaches provide means for broadening abductive logic of discovery.

How to conceptualize the area of discovery?

For Peirce, one basic motivation for abduction was his contention that there has to be a way by which people have managed so successfully to find fertile and true theories for various phenomena. According to Peirce, this could not have been a random process, because there would not have been enough time for that in human history (CP 7.220, 1901; also CP 5.172, 1903; CP 5.591, 1903). Both an inferential and an instinctual interpretation of abduction can be developed so that they clarify how abduction makes this success understandable (Paavola & Hakkarainen 2005). But they are not enough in themselves.

Hanson maintained that there are reasons for suggesting hypotheses and not just for accepting and testing them, against the prevalent thought style in philosophy of science in his time (Hanson 1961). What he meant with "logic of discovery" was a conceptual inquiry concerning the area of discovery, not a manual or an algorithm for making discoveries (ibid., 21). Hanson wanted to conceptualize "the dawning of an explanation" and the way that scientists produce, or "press on" for new explanatory hypotheses on the basis of data and phenomena, and he took Peirce's abduction as a central conceptual means for doing this (Hanson 1958, 85-90). Later, when discovery was recovered as an important topic for philosophy of science, Hanson was praised as an important forerunner for discovery, but his formulations of abduction were treated as insufficient for explaining the area of discovery. The basic criticism against Hanson was that his formulations of abduction failed to provide a basis for a logic of discovery, and gave a too narrow picture of the area of discovery (see Nickles 1980, 22-25). Similar doubts still linger on discussions around abduction and a logic of discovery. There are doubts if it is even in principle possible to develop a separate

"logic of discovery", and specific formulations of abductive inference have been criticized as inadequate (Kapitan 1990; 1992).

I maintain that a Hansonian research programme concerning logic of discovery can be continued although his own formulations of abduction are not entirely adequate. Hanson was not so much searching for a separate logic of discovery, but maintaining that the way that scientists press on for new explanations and hypotheses can be conceptualized. Discovery is not just about logic or reasoning, but still there *are* also reasons for suggesting hypotheses, and various forms of abduction are one way of conceptualizing these processes.

For me, the gist of abduction is those things which constrain and guide the search for new, tentative ideas and hypotheses. Abduction is a weak form of inference where clue-like signs, tentative restrictions, and search for overarching patterns are used as aids when new ideas are searched for. Although it is a weak form of inference, it is better than pure chance. There is no rule-book for making discoveries, but still, abductive processes help us to understand how new ideas are searched for.

According to prevalent creativity myths in our culture, new ideas burst out from the minds of lone geniuses fighting against social conventions and traditional, stagnant thought styles of their time (Sawyer 2006, 259). Or they are the result of reasoning powers of these exceptional people. These myths are not totally wrong. New ideas often require that fossilized ways of interpreting things be seen in a new light, and often individuals start these processes by pushing forward new lines of research and new, insightful ideas. But this is just one part of the story. Individuals work in particular historical and cultural contexts and in an interaction with other people and their environment while developing new ideas. Here the ideas of distributed cognition become important.

Abduction and distributed cognition

Distributed cognition is an approach within cognitive sciences which emphasizes that human cognition is not confined to individuals and within individuals' minds but is distributed in essential ways to surrounding physi-

cal, social, and cultural environments and to long-term temporal processes. It emphasizes social and situational aspects of cognition rather than mental constructions within human mind. The term itself has become popular since the late 1980s (see Oatley 1991), but the idea has many predecessors, such as L. S. Vygotsky's idea of the mediated nature of human cognition (Vygotsky 1978).

Basic interpretations of abduction seem to be in contrast to distributed cognition, or at least they do not promote that kind of thinking. If abduction is interpreted inferentially (or syllogistically), it means that the process is something a-personal (a relationship between premises and a conclusion), or done by an individual reasoner within his or her mind. If abduction is interpreted instinctually or close to perception, it seems to be a process where social interaction, or cultural and material artefacts do not have a prominent role. Much, of course, depends on how Peirce's notion of an instinct (or perception) is interpreted. There are elements also within inferential and instinctual interpretations of abduction which have connections to the ideas of distributed cognition. Abduction starts from an interaction between a human inquirer and its environment (see Nesher 2001). A starting point for abduction is perceptual judgments and surprising phenomena which instigate the search for new ideas (CP 5.182-194, 1903). Abduction is a part of an interrogative process (CP 6.525, 1903; CP 6.469, 1908; Sintonen 2004; Hookway 2005; Paavola, Hakkarainen & Sintonen 2006). Peirce also held that abduction is closely related to, or even the same as, the *maxim of pragmatism* (or "pragmatic maxim"), that is, explanatory hypotheses produced should be capable of experimental verification (CP 5.195-198, 1903).

Peirce's philosophy in general has affinities with the ideas of distributed cognition. According to Peirce's semiotic perspective and in opposition to mentalistic tradition, even individual minds are very fundamentally social, developing in intersubjective sign processes (Colapietro 1989). Peirce's anti-Cartesianism meant that signs mediate human cognition in a fundamental sense (CP 5.213-357, 1868-1869). For Peirce, signs include all kinds of sign processes, not just, for example, symbolic signs but also indexical signs, some of which have material form, where the real connection of signs with

dynamic objects is central (see Liszka 1996). Peter Skagestad (1993; see also Magnani 2006) has also shown how Peirce represented an "augmentationist" approach to mind and knowledge, according to which human thinking should not be seen so much as a process that happens inside the human head but as something where human mind uses various sorts of external artefacts, tools, and methodologies. For Peirce, the augmenting things were especially signs and sign processes.

Peirce did not, however, emphasize social, cultural, or material elements in relation to *abduction* in particular. Karl-Otto Apel has remarked that although Peirce emphasized the role of community especially in his early works concerning human cognition, he seemed to interpret abduction as if it were a "logically mediated exchange of information between each individual person and nature" (Apel 1981, 170-171). What is missing is the role of social interaction and community of interpretation as a basis for new ideas. Peirce emphasized the guessing instinct, or "*il lume naturale*" (a sort of a natural bent) in his later writings (CP 1.80-81, c. 1896), and neglected "*il lume culturale*" (Bonfantini & Proni 1983, 134; Bonfantini 1988, 1253-1254). Peirce did not emphasize material objects or external artefacts in setting out a basis for abduction (also "*il lume materiale*" was then neglected). Similarly N. R. Hanson mostly neglected the role of material and social elements in his analysis of discovery.

There are some recent analyses which interpret abduction within distributed cognition. Keith Oatley has made this connection explicitly (Oatley 1996). When scientists are developing new ideas they are "standing on the shoulders of giants", that is, they are using cultural knowledge produced by other researches (ibid., 137). Inquiry is basically a social process: "[w]hile one person puts forward a hypothesis, someone else sees what is wrong with it, yet others find syntheses and provide wider contexts" (Oatley 1991, 102). Even when the inquirer is apparently doing his research individually, he or she is using and developing concepts, theories and observations developed by others, and trying to communicate, develop and defend them for other people. So the focus should not be just on an individual having insights or

making inferences but also on the interaction with social and material environment which provide a basis for these insights and inferences.

Lorenzo Magnani has emphatically developed ideas belonging to distributed cognition especially with the notion of *manipulative abduction* (Magnani 2001, 53-69; 2004). Manipulative abduction "happens when we are thinking *through* doing and not only, in a pragmatic sense about doing" (Magnani 2004, 229). It uses "epistemic mediators" or external forms of representations (like diagrams and models) when people are collaboratively solving some problems or performing tasks. It has its basis in both concrete manipulations of external objects and on cognitive division of labour among people. Models of cognition should take into account the interplay between internal and external representations (Magnani 2006). Human beings are able to use various kinds of representations, which are especially important when something new is created.

How do the ideas of distributed cognition change the interpretation of abductivist methodology? Roy Pea (1993, 65-67) has maintained that prevalent "myths" of standard problem-solving models present inquiry in linear stages and as constructions done by individuals with mental means; whereas according to distributed cognition problem solving is more about cyclic, iterative systems done by social communities with external representations and artefacts and in relation to features of the environment. If abduction is presented solely as a first stage of inquiry where a hypothesis is derived on the basis of premises, or on the basis of an instinct (or an insight), it leaves this larger framework very much out of sight. From this point of view the problem of inferential and instinctual interpretations is not so much that they are wrong but they focus too narrowly on some aspects of the overall process of discovery.

Distributed cognition gives extra means for understanding those elements which guide and constrain abductive search for new and plausible ideas. It emphasizes, for example, the points below. It is not being claimed that these are, necessarily, absent in Peirce, especially as regards 'seed' form; they are laid out to provide a basis for detailed consideration of possible

elaboration and extension of Peirce's framework to include distributed cognition.

1) The search for abductive ideas has its basis in long-term, iterative and reflective processes. Basic models of abduction give an impression that new hypotheses turn up by starting from a surprising phenomenon and finding a hypothesis candidate to explain that phenomenon (if that hypothesis turns out not to be successful, a new candidate must be produced, and so the process goes on). An instinctual alternative is to say that new ideas turn out as sudden moments of insight when the inquirer realizes a new pattern which explains previously puzzling data (Hanson 1958). Both of these approaches leave out phases preceding these. Ideas are often worked out for a long period of time. They develop and change during this process, but new ideas have their bases in the phases that help to see the constraints and clues that provide guidance to the discovery process. As an example, Hanson (1958, 72, 89) has mentioned that it took, in all, 34 years for Galileo Galilei to develop his constant acceleration hypothesis, and abductive methodology should clarify this whole process, not just the end result or the crucial insight. The same applies to all major discoveries[3]. The elements and material for new ideas develop during a long process and in relation to hypotheses and ideas tried out before. The abductive process is temporally distributed.

2) In empirical sciences clues and elements instigating and constraining new ideas are not just propositional but come from an interaction between the inquirer(s) and the object of research.[4] In a paper called "Guessing" Peirce maintained that human beings are able to notice "intimations of truth" (CP 7.46, 1907) as an aid when

3 And probably this applies also to minor, everyday discoveries. If I, for example, invent an original and a good idea of what to buy for a friend as a birthday present, it has been formed not just by a sudden insight or one line of an inference, but by various previous constraints and clues concerning, my knowledge, e.g., those things that are usually bought as presents, what kind of things my friend likes, and so on. In order to understand properly how the new idea comes up, these preparatory processes should be taken into account.

4 Cf. also a "multimodal" approach on abduction, that is, Thagard has emphasized that for human problem solvers clues and hypotheses within abduction can also be other than propositional or verbal representations, like sights, sounds, smells, touches, emotional feelings, visual patterns, and kinesthetic experiences (Thagard, forthcoming). Thagard connects this to neural mechanism and to a neurocomputational theory of abductive inference, but it can be extended to cover also external representations (Magnani 2006).

guessing new explanations for intricate problems (Peirce 1929)[5]. Peirce did not speculate much what is the source of these "intimations" or clues for new ideas. From the point of view of distributed cognition, one clear source is the human encounter with the objects and events of the material world. The abductive search is then materially distributed.

3) Not just propositional means but also social practices guide human inquiry. Research practices and paradigmatic ways of doing research in a particular field in a particular historical situation on the basis of its historical background very much determine how things are understood and developed. It may be argued that the discovery of new ideas and theories means that these old habits of action are rejected or radically changed. Still, not all practices and habits can be changed at once, but they are partly guiding the process of inquiry. Thus it can be said that cognition is socio-practically distributed.

4) Ideas are not developed just by individuals, but human beings build on each other's ideas in a fundamental sense. Human beings are social and communicative animals. Other people give elements and clues for new ideas (both consciously and unconsciously) when commenting, discussing, or writing on each other's ideas. In this sense, a basic unit of an analysis for understanding processes of discovery should rather be scientists as members of research communities, or networks of inquirers than an individual scientist doing research. Inquiry is socially distributed.

5) Inquirers use culturally shared knowledge in developing new ideas. The timescale of human activity reaches beyond individuals. Culturally created and shared knowledge, artefacts, methods, and tools provide means for human inquiry. Inquirers do not need to start from scratch because these cultural resources guide the inquiry. Very often to develop something new means that culturally stabilized ways of doing or interpreting things are changed or neglected, but not all can be changed or neglected at once. Inquiry is also culturally distributed.

5 Some parts of this paper are published in Peirce CP 7.36-48, 1907. The whole paper is published in *Hound & Horn* magazine from 1929, pp. 267-282, and described at length in Sebeok & Umiker-Sebeok 1983.

6) Not just inferences within the inquirer's "head" are important, but also the meaning of various external tools and artefacts. External artefacts, tools and representations are central when ideas are socially and culturally developed and dispersed. External, shared objects and artefacts are important even when inquirers have competing theories about their meaning (Magnani 2006). People externalize their ideas to external representations because otherwise they would not be available for others to use. Inquirers use, modify and create these external artefacts while developing new ideas. Ideas are distributed in external artefacts.

Distributed cognition challenges previous models concerning inquiry, knowledge and thought. It gives new means for conceptualizing processes of discovery. Instead of concentrating on processes within one's mind, or on an interaction between the inquirer and nature, or between people it means that new kinds of conceptualizations concerning human activity must be developed. It can be maintained that its basic form is a "three-player" situation, that is, an interaction with the inquirer, a community of fellow interpreters, and "nature" (i.e., the object of inquiry) (see Pera 1994; cf. Davidson 2001).[6]

Conclusion

The concept of distributed cognition gives new means for conceptualizing abductive processes of discovery. Peirce's writings on abduction have led to emphasize inferential or "instinctual" interpretations of abduction, but Peirce's texts are amenable to the ideas emphasized with distributed cognition. Peirce's overall philosophy is *not* hostile to the ideas of distributed cognition, but much research remains to be done on what it means that abductive processes are socially, culturally, materially and temporally

6 On the basis of this, we have been developing a "trialogical" model of inquiry conceptualizing an approach where inquirers are developing collaboratively and with cultural means in long-term processes some shared objects of activity (Paavola, Hakkarainen & Sintonen 2006). The term "trialogue" (cf. a bit different but related use of the term by Wiley 1994) is a contrast to "monological" models of human cognition (concentrating on processes within one's mind) and to "dialogical" models (concentrating on an interaction between people or between a human being and nature).

distributed. N. R. Hanson brought forward emphatically the idea of logic of discovery, that is, the need to conceptualize the area of discovery. I maintain that Hanson's research programme can be expanded to areas of distributed cognition. I think that in order to understand better the area of discovery, the aim is not to *substitute* prevalent interpretations of abduction as such, but to supplement and extend these models through a broader perspective on abduction. Understanding processes of discovery requires that various layers of abduction be taken into account.

Acknowledgements

If it would sound reasonable, I would like to thank all those researchers who have written and shall write on abduction, and all those situations which have provided material for abduction. But before that, I wish to thank Hal White for editing and giving pointed comments on the paper, and especially the organizers and all the participants of the "Abduction and the Process of Scientific Discovery" meeting at Lisbon (May 2006) of arranging such a homely and exciting conference with so many abductive details.

References

Anderson, Douglas R. (1987). *Creativity and the Philosophy of C.S. Peirce*. Dordrecht: Martinus Nijhoff Publishers.

Apel, Karl-Otto (1981). *Charles S. Peirce: From pragmatism to pragmaticism*. Amherest, MA: University of Massachusetts Press.

Bonfantini, Massimo A. (1988). "Paper Introducing the Panel on Abduction: Four Questions on Abduction", in Michael Herzfeld and Lucio Melazzo (Eds.) *Semiotic Theory and Practice. Proceedings of the Third International Congress of the IASS Palermo, 1984*. Berlin, New York, Amsterdam: Mouton de Gruyter: 1245-1257.

Bonfantini, Massimo A. & Proni, Giampaolo (1983). "To Guess or Not to Guess?", in Umberto Eco & Thomas A. Sebeok (Eds.) (1983). *The Sign of Three. Dupin, Holmes, Peirce*. Bloomington: Indiana University Press, 119-134.

Burks, Arthur W. (1946). "Peirce's Theory of Abduction", *Philosophy of Science* 13, 301-306.

Colapietro, Vincent (1989). *Peirce's Approach to the Self: A Semiotic Perspective on Human Subjectivity*. Albany: State University of New York.

Davidson, Donald (2001). *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

Eco, Umberto and Sebeok, Thomas A. (Eds.) (1983). *The Sign of Three. Dupin, Holmes, Peirce*. Bloomington: Indiana University Press.

Fann, K. T. (1970). *Peirce´s Theory of Abduction*. The Hague: Martinus Nijhoff.

Frankfurt, H. G. (1958). Peirce's account of inquiry, *The Journal of Philosophy* 55(14): 588-592.

Hanson, Norwood Russell (1958). *Patterns of Discovery*. Cambridge: University Press.

Hanson, Norwood Russell (1961). "Is there a logic of scientific discovery?", in Herbert Feigl & Grover Maxwell (Eds.), *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart and Winston, 20-35.

Hookway, Christopher (2005). "Interrogatives and uncontrollable abductions", *Semiotica* 153(1/4): 101-115.

Hutchins, Edwin (1995). *Cognition in the Wild*. Cambridge, MA: MIT.

Kapitan, Tomis (1990). "In What Way is Abductive Inference Creative", *Transactions of the Charles S. Peirce Society* 26(4): 499-512.

Kapitan, Tomis (1992). "Peirce and the autonomy of abductive reasoning", *Erkenntnis* 37: 1-26.

Liszka, James J. (1996). *A General Introduction to the Semeiotic of Charles Sanders Peirce*. Bloomington and Indianapolis: Indiana University Press.

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

Magnani, Lorenzo (2004). "Model-based and manipulative abduction in science", *Foundations of Science* 9(3): 219-247.

Magnani, Lorenzo (2006). "Multimodal Abduction. External Semiotic Anchors and Hybrid Representations", *Logic Journal of the IGPL* 14(2): 107-136.

Nesher, Dan (2001). "Peircean Epistemology of Learning and the Function of Abduction as the Logic of Discovery", *Transactions of the Charles S. Peirce Society* 37(1): 23-57.

Nickles, Thomas (1980). "Introductory essay: scientific discovery and the future of philosophy of science", in Nickles, T. (Ed.), *Scientific Discovery, Logic, and Rationality*. Dordrecht: D.Reidel Publishing Company.

Oatley, Keith (1991). "Distributed cognition", in H. Eysenck, A. Ellis, E. Hunt, & P. Johnson-Laird (Eds.) *The Blackwell dictionary of cognitive psychology*. Oxford: Blackwell Reference, pp. 102-107.

Oatley, Keith (1996), "Inference in narrative and science", in David R. Olson & Nancy Torrance (Eds.), *Modes of thought. Exploration in culture and cognition*. Cambridge: Cambridge university press.

Paavola, Sami (2005). "Peircean abduction: instinct, or inference?", *Semiotica* 153(1/4): 131-154.

Paavola, Sami & Hakkarainen, Kai (2005). "Three Abductive Solutions to the Meno Paradox – with Instinct, Inference, and Distributed Cognition", *Studies in Philosophy and Education* 24(3-4): 235-253.

Paavola, Sami, Hakkarainen, Kai, & Sintonen, Matti (2006). Abduction with Dialogical and Trialogical Means. *Logic Journal of the IGPL* 14(2), 137-150.

Pea, Roy D. (1993). "Practices of distributed intelligence and designs for education", in G. Salomon (Ed.) *Distributed Cognitions: Psychological and Educational Considerations*. Cambridge: Cambridge University Press, pp. 47–87.

Peirce, Charles S. (1929). "Guessing", *Hound & Horn* 2(3): 267-282.

Peirce, Charles S. (1931-1958). *Collected Papers of Charles Sanders Peirce*, vols. 1 6, Hartshorne, C. and Weiss, P., eds.; vols. 7-8, Burks, A. W., ed. Cambridge, Mass: Harvard University Press.

Pera, Marcello (1994). *The Discourses of Science*. Chicago: The University of Chicago Press,.

Salomon, Gavriel (Ed.) (1993). *Distributed Cognitions: Psychological and Educational Considerations*. Cambridge: Cambridge University Press.

Sawyer, Keith, R. (2006). *Explaining Creativity. The Science of Human Innovation*. Oxford: Oxford University Press.

Sebeok, Thomas A. & Umiker-Sebeok, Jean (1983). "You Know My Method": A Juxtaposition of Charles S. Peirce and Sherlock Holmes", in Umberto Eco & Thomas A. Sebeok (Eds.), The Sign of Three Dupin, Holmes, Peirce. USA: Indiana University Press, pp. 11-54.

Shanahan, Timothy (1986). "The First Moment of Scientific Inquiry: C. S. Peirce on the Logic of Abduction", *Transactions of the Charles S. Peirce Society* 22(4): 450-466.

Sintonen, Matti (2004). "Reasoning to hypotheses: Where do questions come?", *Foundations of Science* 9(3): 249 – 266.

Skagestad, Peter (1993). "Thinking with machines: Intelligence augmentation, evolutionary epistemology, and semiotic", *The Journal of Social and Evolutionary Systems* 16(2): 157-180.

Thagard, Paul (1981). "Peirce on Hypothesis and Abduction", in Kenneth L. Ketner, et al. (Eds.), *Proceedings of the C. S. Peirce Bicentennial International Congress*. Lubbock: Texas Tech Press, pp. 271-274.

Thagard, Paul (forthcoming). "Abductive inference: From philosophical analysis to neural mechanisms. In A. Feeney & E. Heit (Eds.) *Inductive reasoning: Cognitive, mathematical, and neuroscientific approaches*. Cambridge: Cambridge University Press.

Vygotsky, Lev S. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press.

Wiley, Norbert (1994). *The Semiotic Self*. Cambridge, UK: Polity Press.

2

PROCESSES OF SCIENTIFIC DISCOVERY

Prospective Updating of Theories with Preferences

Luís Moniz Pereira
Centro de Inteligência Artificial, Universidade Nova de Lisboa
lmp@di.fct.unl.pt
Pierangelo Dell'Acqua
Department of Science and Technology, Linköping University
pier@itn.liu.se
Gonçalo Lopes
Centro de Inteligência Artificial, Universidade Nova de Lisboa
goncaloclopes@gmail.com

1. Introduction

Logic program semantics and procedures have been used to characterize preferences among the rules of a theory [5]. Whereas the combination of such rule preferences with program updates and the updating of the preference rules themselves [2] have been tackled, a crucial ingredient had been missing, namely the consideration of abductive extensions to a theory, and the integration of revisable preferences among such extensions. The derivations from this preferential theory revision [6] to prospective reasoning [12] and model-based diagnosis are the main subject of this paper.

We take a theory expressed as a logic program under the stable model semantics, already infused with pairwise conditional preferences amongst rules. The possible abductive extensions to the theory can be expressed by a set of abducible literals, over which we should also be able to establish conditional priority relations.

First we supply some epistemological background to the problem at hand, and describe the general requirements of model-based diagnosis and its relation to abductive reasoning. We then consider a real-world example

of abductive diagnosis in medical reasoning and proceed to explain the importance of the proposed approach to prospective logic programming. Then we introduce our preferential abduction framework, in order to address the issues raised by the medical example in a context of updatable abductive theories. Some additional examples of preference updating are presented and we conclude with general epistemic remarks about the entire approach and future directions of study.

2. Preferences, Rationality, Theory Revision and AI

The theoretical notions of preference and rationality with which we are most familiar are those of the economists'. Economic preference is a comparative choice between alternative outcomes, whereby a rational (economic) agent is one whose expressed preferences over a set of outcomes exhibits the structure of a complete pre-order.

However, preferences themselves may change. Viewing this phenomenon as a comparative choice, however, entails that there are meta-level preferences whose outcomes are various preference rankings of beliefs, and that an agent chooses a change in preference based upon a comparative pairwise choice between the class of first-order preferences [7].

But this is an unlikely model of actual change in preference, since we often evaluate changes -- including whether to abandon a change in preference -- based upon items we learn after a change in preference is made. Hence, a realistic model of preference change will not be one that is couched exclusively in decision theoretic terms.

Rather, when a conflict occurs in updating beliefs by new information, the possible items for revision should include both the set of conflicting beliefs and a reified preference relation underlying the belief set. The reason for adopting this strategy is that we do not know, a priori, what is more important -- our data or our theory.

Rather, as Isaac Levi has long advocated [11], rational inquiry is guided by pragmatic considerations, not a priori constraints on rational belief. On Levi's view, all justification for change in belief is pragmatic in the sense that justifications for belief fixation and change are rooted in strategies for

promoting the goals of a given inquiry. Setting these parameters for a particular inquiry fixes the theoretical constraints for the inquiring agent. The important point to stress here is that there is no conflict between theoretical and practical reasoning on Levi's approach, since the prescriptions of Levi's theory are not derived from minimal principles of rational consistency or coherence.

Suppose your scientific theory predicts an observation, o, but you in fact observe ¬o. The problem of carrying out a principled revision of your theory in light of the observation ¬o is surprisingly difficult. One issue that must be confronted is what the principle objects of change are. If theories are simply represented as sets of sentences, and prediction is represented by material implication, then we are confronted with Duhem's Problem [8]: *If a theory entails an observation for which we have disconfirming evidence, logic alone won't tell you which among the conjunction of accepted hypotheses to change in order to restore consistency.* The serious issue raised by Duhem's problem is whether disconfirming evidence targets, in a principled manner, the items of a theory in need of revision.

The AGM conception of belief change differs to Duhem's conception of the problem in important respects. First, whereas the item of change on Duhem's account is a set of sentences, the item of change on the AGM conception is a belief state, represented as a pair consisting of a logically closed set of sentences (a belief set) and of selection function. Second, the revised resulting theories are not explicitly represented, when replacing Duhem's account by the AGM postulates approach. What remains in common is what Hansson [10] has called the input-assimilating model of revision, whereby the object of change is a set of sentences; the input item is a particular sentence; and the output is a new set of sentences.

One insight to emerge is that the input objects for change may not be single sentences, but a sentence-measure pair [13], where the value of the measure represents the entrenchment of the sentence and thereby encodes the ranking of this sentence in the replacement belief set [13, 17, 18]. But once we acknowledge that items of change are not belief *simpliciter*, but belief and order coordinates, then there are two potential items for change:

the acceptance or rejection of a belief and the change of that belief in the ordering. Hence, implicitly, the problem of preference change appears here as well. Within the AGM model of belief change, belief states are the principal objects of change: propositional theory (belief set) changed according to the input-assimilating model, whereby the object of change (a belief set) is exposed to an input (a sentence) and yields a new belief set.

Computer science has adopted logic as its general foundational tool, while Artificial Intelligence (AI) has made viable the proposition of turning logic into a bona fide computer programming language. AI has developed logic beyond the confines of monotonic cumulativity, typical of the precise, complete, endurable, condensed, and closed mathematical domains, in order to open it up to the non-monotonic real world domain of imprecise, incomplete, contradictory, arguable, revisable, distributed, and evolving knowledge. In short, AI has added dynamics to erstwhile statics. Indeed, classical logic has been developed to study well-defined, consistent, and unchanging mathematical objects. It thereby acquired a static character.

AI needs to deal with knowledge in flux, and less than perfect conditions, by means of more dynamic forms of logic. Too many things can go wrong in an open non-mathematical world, some of which we don't even suspect. In the real world, any setting is too complex already for us to define exhaustively each time. We have to allow for unforeseen exceptions to occur, based on new incoming information. Thus, instead of having to make sure or prove that some condition is not present, we may assume it is not (the Closed World Assumption - CWA), on condition that we are prepared to accept subsequent information to the contrary, i.e. we may assume a more general rule than warranted, but must henceforth be prepared to deal with arising exceptions.

Much of this has been the focus of research in logic programming. This is a field which uses logic directly as a programming language, and provides specific implementation methods and efficient working systems to do so. Logic programming is, moreover, much used as a staple implementation vehicle for logic approaches to AI.

3. Model-Based Diagnosis and Future Prospecting

Abduction can also have an important rule in model-based diagnosis. In diagnosis, we are concerned with developing precise methodologies and techniques that can determine whether the behaviour of a system (or artefact) is correct. The artefact can in itself be a knowledge theory, in which case we are trying to ascertain whether the conclusions inferred from the theory are valid.

If the system is not functioning correctly, the diagnosis technique should be able to determine, as accurately as possible, which part of the system is failing, and the kind of fault it is facing. The process of diagnosis is usually based on observations which provide information on the current behaviour of the system.

Model-based diagnosis is an example of abductive reasoning using a model of the system that is to be diagnosed. This model describes the behaviour of the system, or artefact. The model can itself be the subject of diagnosis. This model is usually an abstraction of the behaviour of the system and can be incomplete. The faulty behaviour may also be unknown (e.g. no fault model is represented). Given the set of observations about the system, the diagnoser simulates the system using the model, and compares the observations actually made to the observations predicted by the simulation.

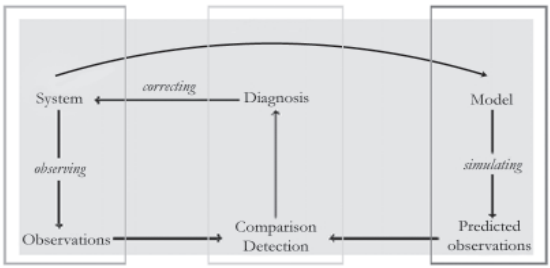


Figure 1: The three components of model-based diagnosis.

Construction of the model can be expressed by rules indicating those cases in which the behaviour of the system is normal or abnormal:

$Abnormal(S) \Rightarrow InternalBehaviour(Int1), Observable(Obs1).$

$\neg Abnormal(S) \Rightarrow InternalBehaviour(Int2), Observable(Obs2).$

Given the set of observations *Obs*, the problem is to determine whether the system behaviour is normal or not (i.e. ¬*Abnormal(S)* or *Abnormal*(*S*)). A system is said to be diagnosable if we are able to determine, without ambiguity, a single diagnosis, independently of the behaviour it exhibits.

However, since the model of our system can be incomplete, this means that there may exist faults and behaviours unaccounted for by the model. Abductive reasoning in the context of evolving logic programs [3, 6, 12] can help us extend the theory to account for unforeseen situations, using expectations to generate the possible scenarios for diagnosis.

This involves a notion of simulation, in which the diagnoser is capable of conjuring up hypothetical *what-if* scenarios and formulating abductive explanations for both external and internal observations. Since we may have multiple possible scenarios to choose from, we need some form of preference specification, which can be either a priori or a posteriori. A priori preferences are embedded in the knowledge representation theory using the framework presented below and can be used to produce the most relevant hypothetical abductions for a given state and observations, in order to conjecture possible future states. A posteriori preferences represent choice mechanisms, which enable the program to commit to one of the hypothetical scenarios engendered by the relevant abductive theories. These mechanisms may trigger additional simulations in order to posit which new information to acquire, so more informed choices can be enacted, in particular by restricting and committing to some of the abductive explanations along the way.

3.1. Differential Medical Diagnosis

In medicine, differential diagnosis is the systematic method physicians use to identify the disease causing a patient's symptoms. Before a medical condition can be treated it must first be correctly diagnosed. The physician begins by observing the patient's symptoms, taking into consideration the patient's personal and family medical history and performing additional examinations if current information is lacking. Then the physician lists the most likely causes. The physician asks questions and performs tests to eliminate possibilities until he or she is satisfied that the single most likely cause

has been identified. The term differential diagnosis also refers to medical information specially organized to aid in diagnosis, particularly a list of the most common causes of a given symptom, annotated with advice on how to narrow down the list.

This listing of the most likely causes is clearly an abduction process, supported by initial observations. Encoding this process in logic programming would result in a set of literals and rules representing both the expected causes and the knowledge representation theory leading up to them. This set could be subsequently refined through a method of prospection with a priori and a posteriori preferences, like the one described above. This prospective process can result in the decision to perform additional observations on the external environment, and be iterated.

In this case, the external observations correspond to the disease's signs and other examinations performed by the physician during the process of diagnosis. The abductive explanations correspond to the possible causes (e.g. medical conditions or simply bad habits) responsible for the symptoms. A priori preferences and expectations are necessarily determined according to the patient's symptoms and they can be updated during the course of the prospection mechanism. An appropriate knowledge representation theory can be derived from current medical knowledge about the causes of a given disease.

It being such a good example of abduction, we provide further on a practical example of iterated differential diagnosis, based upon real medical knowledge from the field of dentistry, in order to better illustrate the capabilities of abductive reasoning systems. An encoding of the problem in the ACORDA[1] system will then be proffered, along with results from an interactive diagnosis session.

1 ACORDA is a self-evolving prospective logic programming system which uses abduction to generate its own possible future scenarios, using a theory of preferences among these hypotheses to guide the process of finding the best explanations for present observations. For more details concerning its implementation see [12].

4. Framework

4.1. Language

Let L be a first order language. A domain literal in L is a domain atom A or its default negation *not* A, the latter expressing that the atom is false by default (CWA). A domain rule in L is a rule of the form:

$$A \leftarrow L_1,...,L_t \ (t \geq 0)$$

where A is a domain atom and $L_1,...,L_t$ are domain literals. To express preference information, L contains priority rules. Let $N = \{n_{r1},...,n_{rk}\}$ be a name set containing a unique name for every domain rule in L. Given a domain rule r, we write n_r to indicate its name. A priority atom is an atom of the form $n_r < n_u$, where $\{n_r,n_u\} \in N$. $n_r < n_u$ means that rule r is preferred to rule u. We assume that names in N do not include "<" itself. A priority rule in L is a rule of the form:

$$n_r < n_u \leftarrow L_1,...,L_t \ (t \geq 0)$$

where $n_r < n_u$ is a priority atom and every L_i $(1 \leq i \leq t)$ is a domain literal or a priority literal.

We use the following convention. Given a rule r of the form $L_0 \leftarrow L_1,...,L_t$, we write $H(r)$ to indicate

L_0, $B(r)$ to indicate the conjunction $L_1,...,L_t$. We write $B^+(r)$ to indicate the conjunction of all positive literals in $B(r)$, and $B^-(r)$ to indicate the conjunction of all negated literals in $B(r)$.When $t = 0$ we write the rule r simply as L_0.

A (logic) program P over L is a finite set of domain rules and priority rules. Every program P has associated a set of *abducibles* A_P, without rules in P. Abducibles may be thought of as hypotheses that can be used to extend the current theory of the agent, in order to provide hypothetical solutions or possible explanations for given queries.

A 2-valued interpretation M of L is any set of literals from L that satisfies the condition that, for any atom A, precisely one of the literals A or *not* A belongs to M. We say that an interpretation M satisfies a conjunction of literals $L_1,...,L_t$, if every literal L_i in the conjunction belongs to M.

4.2. Declarative Semantics

In the remainder of this section we let P be a program over L, A_p the set of abducibles of P, and M

an interpretation of L. We write *least*(P) to indicate the least model of P. We adopt the following two definitions from [9], and Definitions 4 and 5 from [2], and refer the reader to those sources for a detailed exposition.

Definition 1. *The set of default assumptions of P with respect to M is:*

$$Default(P,M) = \{not\ A : \neg\exists r \in P \text{ such that } H(r) = A \text{ and } M \models B(r)\}$$

Definition 2. *M is a stable model of P iff* $M = least(P \cup Default(P,M))$.

Definition 3. *Let $\Delta \subseteq A_p$, M is an abductive stable model with hypotheses Δ of P iff:*

$$M = least(P^+ \cup Default(P^+,M)), \text{ where } P^+ = P \cup \Delta.$$

Note that the abducibles in A_p are defined false by default whenever they are not abduced. Given a program P, to compute which of its abductive stable models are preferred according to the priority relation $<$, we remove (from the program) all the unsupported rules together with the less preferred rules defeated by the head of some more preferred one, in a priority rule. Unsupported rules are those whose head is true in the model but whose body is defeated by the model, i.e. some of its default negated atoms are false in it.

Definition 4. *The set of unsupported rules of P and M is:*

$$Unsup(P,M) = \{r \in P : M \models H(r), M \models B^+(r) \text{ and } M \not\models B^-(r)\}.$$

Definition 5. *$Unpref(P,M)$ is a set of unpreferred rules of P and M iff:*

$$Unpref(P,M) = least(Unsup(P,M) \cup Q), \text{ where}$$

$$Q = \{r \in P : \exists u \in (P - Unpref(P,M)) \text{ such that } M \models n_u < n_r, M \models B^+(u), \text{ and } [\, not\ H(u) \in B^-(r) \text{ or } (not\ H(r) \in B^-(u), M \models B(r))\,]\}.$$

A rule r is unpreferred if it is unsupported or there exists a more preferred rule u (which is not itself unpreferred) such that the positive literals in B(u) hold, and r is defeated by u or r attacks (i.e., attempts to defeat) u. Note

that only domain rules can be unpreferred since it is required that $M \models n_u < n_r$ holds, where n_r and n_u are names of domain rules.

The following definition introduces the notion of preferred abductive stable model. Given a program P and a set Δ of hypotheses, a preferred abductive stable model with hypotheses Δ of P is a stable model of the program that contains all the hypotheses in Δ, and all those rules in P that are not unpreferred.

Definition 6. *Let $\Delta \subseteq A_P$ and M an abductive stable model with hypotheses Δ of P. M is a preferred abductive stable model with hypotheses Δ of P iff:*

1. *if $M \models n_{r1} < n_{r2}$, then $M \not\models n_{r2} < n_{r1}$*
2. *if $M \models n_{r1} < n_{r2}$ and $M \models n_{r2} < n_{r3}$, then $M \models n_{r1} < n_{r3}$*
3. *$M = least(P^+ - Unpref(P^+, M) \cup Default(P^+, M))$, with $P^+ = P \cup \Delta$.*

Conditions 1 and 2 state that the preference relation $<$ is required to be a strict partial order. When the language contains only domain rules and priority rules (that is, there are no abducibles), the semantics reduces to the Preferential semantics of Brewka and Eiter [5]. If integrity constraints are introduced, this semantics generalizes to the Updates and Preferences semantics of Alferes and Pereira [2], which extends updatable logic programs with updatable preferences. Our semantics takes the latter, without formally addressing updating, and complements it with modifiable abducibles.

Definition 7. *An abductive explanation for a query G is any set $\Delta \subseteq A_P$ of hypotheses such that there exists a preferred abductive stable model M with hypotheses Δ of P for which $M \models G$.*

A program may have several abductive explanations for a query G.

5. Preferring Abducibles

The evaluation of alternative explanations is a central problem in abduction, because of the combinatorial explosion of possible explanations to handle. Thus, it is important to generate only the explanations that are relevant for the problem at hand. Several approaches have thus far been

proposed, often based on some global criteria, which has the drawback of generally being domain independent and computationally expensive. An alternative to global criteria for competing alternative assumptions is to allow the theory to contain rules encoding domain specific information about the likelihood that a particular assumption is true.

In our approach, preferences among abducibles can be expressed in order to discard unwanted assumptions. Technically, preferences over alternative abducibles are coded into even cycles over default negation, and preferring one of the rules will break the cycle in favour of one abducible or another. The notion of expectation is employed to express preconditions for assuming abducibles. An abducible can be assumed only if it is confirmed, i.e. there is an expectation for it, and there is not an expectation to the contrary (expect_not).

To express preference criteria among abducibles, we introduce the language L^*. A relevance atom is one of the form $a \lhd b$, where a and b are abducibles. $a \lhd b$ means that the abducible a is more relevant than the abducible b. A relevance rule is a rule of the form:

$$a \lhd b \leftarrow L_1,...,L_t \ (t \geq 0)$$

where $a \lhd b$ is a relevance atom and every L_i $(1 \leq i \leq t)$ is a domain literal or a relevance literal. Let L^* be a language consisting of domain rules and relevance rules.

Example 1. Consider a situation where an agent Claire drinks either tea or coffee (but not both). Suppose that Claire prefers coffee over tea when sleepy. This situation can be represented by a program Q over L^* with set of abducibles $A_Q = \{tea, coffee\}$:

drink ← tea
drink ← coffee
expect(tea)
expect(coffee)
expect_not(coffee) ← blood_pressure_high
coffee ◁ tea ← sleepy

Having the notion of expectation allows one to express the preconditions for an expectation or otherwise about an assumption *a*, and express which possible expectations are *confirmed* (or go through) in a given situation. If the preconditions do not hold, then expectation *a* cannot be confirmed, and therefore *a* will never be assumed. By means of *expect_not* one can express situations where one does not expect something. In this case, when blood pressure is high, coffee will not be confirmed or go through because of the contrary expectation arising as well (and therefore tea will be assumed).

The following definition exploits the relevancy relation $\lhd$ of a program Q to distinguish which of its abductive stable models are relevant.[2]

Definition 8. *Let Q be a program over L^* with set of abducibles A_Q and M an interpretation of L^*. Let $\Delta \subseteq A_Q$. M is a relevant abductive stable model of Q with hypothesis Δ iff:*

1. *for every $x,y \in A_Q$, if $M \models x \lhd y$ then $M \not\models y \lhd x$*
2. *for every $x,y,z \in A_Q$, if $M \models x \lhd y$ and $M \models y \lhd z$, then $M \models x \lhd z$*
3. *$M = least(Q^+ \cup Default(Q^+, M))$, with $Q^+ = Q \cup \Delta$*
4. *Δ is the empty set or a singleton $\Delta = \{a\}$, for some $a \in A_Q$*
5. *if $\Delta = \{a\}$, then $M \models expect(a)$ and $M \not\models expect_not(a)$*
6. *if $\Delta = \{a\}$, then there exists no relevance rule r in Q such that:*
   - *$H(r)$ is $x \lhd a$*
   - *$M \models H(r)$*
   - *$M \models expect(x)$ and $M \not\models expect_not(x)$*

By allowing Δ to be the empty set, or a singleton (condition 4 above) we are stating that the abducibles in A_Q are mutually exclusive. This notion can be generalized to sets of abducibles, by an adequate adaptation of the preference order to one among such sets. It is possible, however, to consider that a set of literals is, in itself, an abducible, so we could even express different interesting combinations of multiple literals for a given problem domain.

2 See [6] for a more detailed exposition and implementation of the abductive stable models semantics in logic programming.

Example 2. Let Q be the program of Example 1. Q has two alternative explanations $\Delta_1 = \{coffee\}$ and $\Delta_2 = \{tea\}$ for the query *drink*. In fact, Q has two relevant abductive stable models:

$M_1 = \{expect(tea), expect(coffee), coffee, drink\}$ with hypotheses Δ_1

$M_2 = \{expect(tea), expect(coffee), tea, drink\}$ with hypotheses Δ_2

for which $M_1 \models drink$ and $M_2 \models drink$. The number of models reduces to one if we add *sleepy* to Q. In this case, *coffee* being an abducible more relevant than *tea* and consequently the only relevant model of Q is $M_1 \cup \{sleepy\}$.

In order to obtain a proof procedure for the language L^*, a syntactical transformation Σ that maps programs over L^* into programs over L has been given in [6], and its soundness demonstrated. The next example illustrates its use.

Example 3. Let Q be the program of Example 1. The transformation Σ maps Q into the program P with abducibles $A_p = \{abduce\}$:

drink ← tea
drink ← coffee

expect(tea)
expect(coffee)

expect_not(coffee) ← blood_pressure_high

coffee ← abduce, not tea, confirm(coffee) (1)
tea ← abduce, not coffee, confirm(tea) (2)

confirm(tea) ← expect(tea), not expect_not(tea)
confirm(coffee) ← expect(coffee), not expect_not(coffee)

1 < 2 ← sleepy

The role of the abducible *abduce* is to enact the assumption of one of the alternative assumptions *tea* or *coffee* needed to prove *drink*. The rules (1) and (2) code the alternative assumptions *tea* and *coffee* into cycles over negation. Rule (1) says that *coffee* can be assumed if *abduce* has been abduced, *tea* is not

assumed, and coffee is confirmed. The last rule in P is a priority rule stating that rule (1) is preferable to rule (2) if *sleepy* holds. P has two preferred abductive stable models with hypotheses Δ = {abduce}:

M_1 = {*abduce*, *confirm*(*tea*), *confirm*(*coffee*), *expect*(*tea*), *expect*(*coffee*), *coffee*, *drink*}

M_2 = {*abduce*, *confirm*(*tea*), *confirm*(*coffee*), *expect*(*tea*), *expect*(*coffee*), *tea*, *drink*}

The number of preferred abductive stable models reduces to one if *sleepy* holds. In that case, the unique preferred abductive stable model would be:

M_3={*abduce, confirm*(*tea*), *confirm*(*coffee*), *expect*(*tea*), *expect*(*coffee*), *coffee*, *drink*, *1 < 2*}

6. Exploratory Data Analysis

Another application of expressing preferences over abducibles is that of *exploratory data analysis*, which aims at suggesting a pattern for further inquiry, and contributes to the conceptual and qualitative understanding of a phenomenon.

Assume that an unexpected phenomenon, x, is observed by an agent Bob, and that Bob has three possible hypotheses (abducibles) a, b, c, capable of explaining it. In exploratory data analysis, after observing some new facts, we abduce explanations and explore them to check predicted values against observations. Though there may be more than one convincing explanation, we abduce only the more plausible of them. The next example illustrates exploratory data analysis.

Example 4. Let the program Q over L^*, with abducibles A_Q = {a,b,c}, be the theory of agent Bob:

$x \leftarrow a$ $x \leftarrow b$ $x \leftarrow c$ *expect*(a) *expect*(b) *expect*(c) $a \lhd c \leftarrow$ *not e* $b \lhd c \leftarrow$ *not e* $b \lhd a \leftarrow d$	meaning: x - the car does not start a - the battery has problems b- the ignition is damaged c - there is no gasoline in the car d - the car's radio works e - Bob's wife has used the car exp - test if the car's radio works.

Q has two relevant abductive stable models capable of explaining observation x:

$M_1 = \{expect(a), expect(b), expect(c), a \lhd c, b \lhd c, a, x\}$ with hypothesis $\Delta_1 = \{a\}$

$M_2 = \{expect(a), expect(b), expect(c), a \lhd c, b \lhd c, b, x\}$ with hypothesis $\Delta_2 = \{b\}$

In this example, we have only a partial relevancy theory over abducibles. Thus, we cannot select exactly one abducible (i.e., one model), as it were the case had we a complete relevancy relation over all abducibles in A_Q. To prefer between a and b, one can perform some experiment exp to obtain confirmation (by observing the environment) about the most plausible hypothesis. To do so, we can employ active rules that are rules of the form:

$$L_1,...,L_t \Rightarrow a : A$$

where $L_1,...,L_t$ are domain literals, and $a : A$ is an action literal. This rule states to update the theory of an agent a with A if its body $L_1,...,L_t$ is satisfied in all relevant abductive stable models. For example, we can add the following rules (where *env* plays the role of the environment) to the theory Q of Bob:

$choose \leftarrow a$

$choose \leftarrow b$

$a \Rightarrow Bob : chosen$

$b \Rightarrow Bob : chosen$

choose $\Rightarrow Bob : (not\ \ chosen \Rightarrow env : exp)$

Initially *Bob* has two hypotheses, a and b, that are capable of explaining the observed phenomena x. Hence, *Bob* must discover the correct one. *Bob* chooses some hypothesis if a or b hold. In this case, *Bob* still has two relevant abductive stable models: $M_3 = M_1 \cup \{choose\}$ and $M_4 = M_2 \cup \{choose\}$. As *choose* holds in both models, the last active rule is triggerable. When triggered, it will add (at the next state) the active rule $not\ \ chosen \Rightarrow env : exp$ to the theory of *Bob*, and, if *not chosen* holds, *Bob* will perform the experiment *exp*. The first two active rules are needed to prevent *Bob* from performing *exp* when *Bob* has chosen one of the abducibles.

7. Modelling Observations

In order to allow specification of model-based diagnosis using the proposed preferential theory framework, it is useful to provide a means to encode observations in the system, both external and internal, respectively, those experiments that the system can perform on the environment, and the observations for which it wants to find a reasonable explanation. The notion of *observable* is used to represent knowledge about such observations.

An observable is a quaternary relation amongst the *observer*, i.e. that which is performing the observation; the *observed*, that which is the target of the observation; the result of the *observation* itself; and the *truth value* associated with the observation. This relation is adequate both to express program based self-triggering of internal queries, as well as observations requested from an exterior oracle to the program, and from the program directly to the environment. For example, the observable

observable(prog,prog,Query,true)

represents an observation in which the observer is the program, that which is observed is also the program, the observation is the specified Query and the positive truth value means the observation must be proven true. In this case, we expect that such an observable, by becoming active via special on_observable rules, triggers the self-evaluation of Query in the current knowledge state, in the process resorting to any relevant and expected abducibles that can account for the observation.

Under certain conditions, these observations can represent not only experiments performed on the environment, but also questions to external agents to acquire additional information with which to prefer among relevant hypotheses. These agents are called *oracles*, and they can be just about anything that is able to interact with the system in order to provide additional information, be it facts or rules. This may represent an instance of learning by being told, with the agent being informed of new pieces of knowledge which allow it to make better choices. The oracle-agent relation can be further complicated in order to include an encoding of levels of trust, or even limited time-spans for the validity of the obtained information, after which

further checks will be performed in order to ascertain whether (or to what extent) the knowledge has proven itself useful.

For now, information derived from oracles can be regarded as an *event*, following the concept introduced in the language for updates LUPS [4]. In the more declarative semantics of updates, EVOLP [3], we can consider an event as an update of a literal or rule which is immediately followed by a deletion of that very same literal or rule.

8. Differential Diagnosis in Dentistry: A Use Case

The process of generating possible explanations to given observations and committing to the resulting choices can be integrated on a larger computational reasoning system, as the one described in Figure 2.

Starting from some state of an evolving knowledge base, we begin by selecting the observations that we would like to explain. We are then able to come up with possible abductive hypotheses that allow us to explain those observations. A priori and a posteriori preferences are used to filter out the most relevant explanatory scenarios, allowing once again the realization of experiments to disambiguate among abducibles.

Considering such an evolving logic system, we are now ready to tackle the process of encoding a real-life situation of medical differential diagnosis using the proposed framework. Consider the following setting from the field of dentistry.

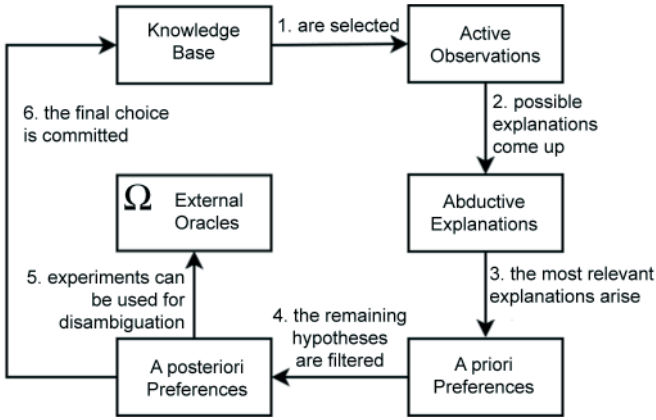


Figure 2: An evolving abductive knowledge base

Example 5. A patient shows up at the dentist with signs of pain upon teeth percussion. The expected causes for the observed signs are:

- Periapical lesion (endodontic or periodontal source)
- Horizontal Fracture of the root and/or crown
- Vertical Fracture of the root and/or crown

Several additional examinations can be conducted to determine the exact cause, namely:

- X-Ray for determination of radiolucency or fracture traces
- X-Ray for determination of periapical lesion source
- Check for tooth mobility
- Measurement of gingival pockets

Aside from presenting multiple hypotheses for diagnosis, the knowledge exhibited by the practitioner must necessarily evolve in time, as he or she performs relevant examinations which will attempt to disqualify all but one of the possible explanations. Current examinations depend on knowledge acquired in the past, which, in turn, will end up influencing the observations and inferences which will be drawn in the future. Developing a system that is capable of modelling the evolution of a program in order to draw such inferences is a demanding but obviously useful challenge.

This setup can be intuitively modelled on top of an abductive logic program:

Initial Signs

on_observable(prog,prog,percussion_pain_cause)←percussion_pain. (A)

First Phase of Differential Diagnosis

percussion_pain_cause ← periapical_lesion.
percussion_pain_cause ← vertical_fracture.
percussion_pain_cause ← horizontal_fracture.

periapical_lesion ← confirm(periapical_lesion).
expect(periapical_lesion) ← profound_caries.

expect(periapical_lesion)←on_observable(prog,prog,percussionpain_cause).(B)
expect_not(periapical_lesion) ← fracture_traces, not radiolucency.

vertical_fracture ← confirm(vertical_fracture).
expect(vertical_fracture) ← on_observable(prog,prog,percussion_pain_cause).
expect_not(vertical_fracture) ← radiolucency, not fracture_traces.

horizontal_fracture ← confirm(horizontal_fracture).
expect(horizontal_fracture) ← on_observable(prog,prog,percussion_pain_cause).
expect_not(horizontal_fracture) ← radiolucency, not fracture_traces.

periapical_lesion ◁ horizontal_fracture ← profound_caries, not percussion_pain.
periapical_lesion ◁ vertical_fracture ← profound_caries, not percussion_pain.

horizontal_fracture ◁ vertical_fracture ← low_mobility.
vertical_fracture ◁ horizontal_fracture ← high_mobility.

Second Phase of Differential Diagnosis

on_observable(prog,prog,periapical_lesion_source)←periapical_lesion. (C)

periapical_lesion_source ← endodontic_lesion.
periapical_lesion_source ← periodontal_lesion.

endodontic_lesion ← confirm(endodontic_lesion).
expect(endodontic_lesion) ← on_observable(prog,prog,periapical_lesion_source).
expect(endodontic_lesion) ← devitalization.
expect(endodontic_lesion) ← not gingival_pockets.
expect_not(endodontic_lesion)←gingival_pockets,notdevitalization. (D)

periodontal_lesion ← confirm(periodontal_lesion).
expect(periodontal_lesion) ← on_observable(prog,prog,periapical_lesion_source).
expect(periodontal_lesion) ← devitalization.
expect(periodontal_lesion) ← gingival_pockets.
expect_not(periodontal_lesion) ← neg_gingival_pockets.
periodontal_lesion ◁ endodontic_lesion ← gingival_pockets.

Available Experiments

radiolucency ← observable(prog,xray,radiolucency).

fracture_traces ← observable(prog,xray,fracture_traces). (E)
high_mobility ← observable(prog,mobility_check,high_mobility).
low_mobility ← observable(prog,mobility_check,low_mobility).
gingival_pockets ← observable(prog,pockets_check,gingival_pockets).
neg_gingival_pockets ← observable(prog,pockets_check,gingival_pockets,false).
devitalization ← observable(prog,periapical_xray,devitalization).

Available Oracles

observable(prog,xray,Q,S) ← oracle,prolog((oracleQuery(xray(Q),T),S = T)).
observable(prog,mobility_check,Q,S) ←
oracle,prolog((oracleQuery(mobility_check(Q),T),S = T)).
observable(prog,pockets_check,Q,S) ←
oracle,prolog((oracleQuery(pockets_check(Q),T),S = T)).
observable(prog,periapical_xray,Q,S) ← oracle,prolog((oracleQuery(periapi
cal_xray(Q),T),S = T)).

General Confirmation Rule

confirm(X) ← expect(X), not expect_not(X).

8.1. Interaction Results

Using the above program as the initial knowledge base, we can provide the patient's signs (i.e. *percussion_pain*) as external updates to the system. Running an abduction cycle over the ACORDA system produces the reasoning steps detailed below. The question marks are used to indicate queries to an external oracle, followed by the answer instances that the oracle generated. Next to each evaluation step we present the number codes relating to the phases from Figure 2.

1. About to launch new introspection on selected active observables:
   [on_observable(prog,prog,percussion_pain_cause)]
2. Relevant abducibles for current introspection:
   [horizontal_fracture,periapical_lesion,vertical_fracture]
3. Partial models remaining after a priori preferences:
   [[horizontal_fracture],[periapical_lesion],[vertical_fracture]]

4. About to launch new introspection on selected active observables:
   [on_observable(prog,prog,percussion_pain_cause)]
5. Confirm observation:
   xray(fracture_traces) (true, false or unknown)? false.
5. Confirm observation:
   xray(radiolucency) (true, false or unknown)? true.
5. Relevant abducibles for current introspection:
   [periapical_lesion]
5. Partial models remaining after a priori preferences:
   [[periapical_lesion]]

The event *percussion_pain* triggers the active observable *percussion_pain_cause* (cf. rule A) which is a program-to-program generated observation, that is, it will cause a top-down query to be launched in order to attempt an explanation of the observable. The system goes down the rule derivation tree for *percussion_pain_cause*, encountering the different possible alternatives derived from the expectation and counter-expectation rules, i.e. the literals that are output in 2. In this case, the expectations are triggered only in the context of the currently active observation (cf. rule B).

Afterwards, the a priori preferences concerning the relevant abducibles come into play in order to filter out less preferred hypothesis. These preferences can themselves be supported by abducibles or further observations. The abductive stable models of the program are then computed, resulting in at least a model for each abducible that could not be defeated by the contextual preference rules.

In our current case, no relevant preferences are present in the model, so the resulting stable models correspond to all the confirmed abducibles: *periapical_lesion*, *horizontal_fracture* and *vertical_fracture*. Since the system was not able to abduce just one model from a priori preferences, it means that the current information is insufficient to provide an adequate informed choice. The prospective search is then relaunched in order to acquire new informa-

tion from the available oracles, in order to search for ways to defeat some of the models.

This means that the oracle mechanisms are activated, and for the new observations the system decrees that experiments can now be performed. A new top-down query is launched, and as the derivation tree is traversed for new confirmation of relevant abducibles, opportunities to perform experiments are encountered. This is the case for the satisfaction of *expect_not*(*periapical_lesion*) which depends on *fracture_traces* and *not radiolucency*.

In an attempt to satisfy *fracture_traces*, the ACORDA encounters an observable clause (cf. rule E). This clause depends on fracture traces being found on X-Ray imaging, so it probes the environment for just such an exam. No fracture traces are found, so periapical lesion can be confirmed under support from the oracle experiment. However, the subsequent experiment for *radiolucency* reveals positive results which can guarantee defeat of the other two abducibles, meaning that *periapical_lesion* is now the only expected and confirmed abducible. ACORDA can now successfully commit to a single diagnosis.

In the cases where the system cannot guarantee a single abductive stable model after exhausting all the introspections, it can query the user to perform the final choice over the list of surviving abducibles, or to provide additional information or experiments to perform in order to continue the diagnosis process. This behaviour could even be refined to launching separate simulations assuming each of the remaining solutions, and performing extra levels of prospective lookahead for additional ways to defeat further models. This addition is considered in future work.

After the abduction of periapical lesion as the most likely cause for the sign of percussion pain on the patient's tooth, the attention of the system turns to satisfy a new observation for the source of such lesion. As such, a new cycle of introspection on top of these results can produce a more detailed diagnosis, as depicted below:

1. About to launch new introspection on selected active observables:
   [on_observable(prog,prog,periapical_lesion_source)]

2. Relevant abducibles for current introspection: [endodontic_ lesion,periodontal_lesion]
3. Partial models remaining after a priori preferences: [[endodontic_lesion],[periodontal_lesion]]
4. About to launch new introspection on selected active observables: [on_observable(prog,prog,periapical_lesion_source)]
5. Confirm observation: pockets_check(gingival_pockets) (true, false or unknown)? true.
5. Confirm observation: periapical_xray(devitalization) (true, false or unknown)? true.
5. Relevant abducibles for current introspection: [endodontic_lesion,periodontal_lesion]
5. Partial models remaining after a priori preferences: [[periodontal_lesion]]

This time, the commitment to the abduction of *periapical_lesion* triggers the activation of program-to-program observable *periapical_lesion_source* (cf. rule C). The relevant expected abducibles that satisfy this observation are *endodontic_lesion* and *periodontal_lesion*, both being confirmed under the current knowledge state. Again, no active preferences hold a priori, that is, there are no preference rules relevant to the current abducibles which do not depend on any external observations. As a result, the Abductive Stable Models of the program correspond once again to all available abducibles.

The choice mechanisms are activated and the top goal is relaunched in an attempt to acquire additional information, activating the oracles for external environment probing. The attempt to defeat *endodontic_lesion* calls for a gingival pockets measurement (cf. rule D), which reveals the existence of pockets in the vicinity of the patient's tooth. On the other hand, a periapical X-Ray is needed to confirm that no endodontic therapy was performed on that tooth. This experiment, however, reveals just that, and so ACORDA is unable to defeat the *endodontic_lesion* abducible using this *expect_not* clause.

Attempting to defeat *periodontal_lesion* by means of *expect_not* clauses proves impossible as well, due once again to the ambiguous results from the experiments. Both abducibles are again confirmed. However, a preference for *periodontal_lesion* is now in place, given the existence of gingival pockets, so a unique Abductive Stable Model emerges, yielding the final diagnosis of *periodontal_lesion*.

The system could now be extended to handle different treatment hypotheses according to the result of the diagnosis. It would just be a matter of adding new triggers for additional observations that would represent the adequate treatment. Abducibles in this case would be the expected treatments and the preference model could be extended to include the patient's own preferences.

9. Revising Relevancy Relations

Relevancy relations are subject to be modified when new information is brought to the knowledge of an individual, or one needs to represent and reason about the simultaneous relevancy relations of several individuals. The resulting relevancy relation may not satisfy the required properties (e.g., a strict partial order - *spo*) and must therefore be revised. We investigate next the problem of revising relevancy relations by means of declarative debugging, via abductive diagnosis.

Example 6. Consider the boolean composition of two relevancy relations, that is, $\lhd = \lhd_1 \cup \lhd_2$. The relevancy relation $\lhd$ may not be a strict partial order, since antisymmetry and transitivity are not necessarily preserved. Consider the following program Q over L^* with abducibles $A_Q = \{a,b,c\}$:

$x \leftarrow a$
$x \leftarrow b$
$x \leftarrow c$
$expect(a)$
$expect(b)$
$expect(c)$

$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_1 \underline{v}$
$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_2 \underline{v}$
$a \lhd_1 b$
$b \lhd_1 c$
$b \lhd_2 a$

where $\underline{u}$ and $\underline{v}$ are variables ranging over the abducibles in A_Q. The program Q does not have any relevant abductive stable model since $\lhd$ is not a strict partial order and therefore conditions 1 and 2 of Definition 8 are not met.

In order to account revising relevancy relations, we introduce the language L^+ extending L^* to contain integrity constraints.

An integrity constraint is a rule of the form:

$$\bot \leftarrow L_1,...,L_t \ (t \geq 0)$$

where $\bot$ is a domain atom denoting contradiction, and $L_1,...,L_t$ are domain or relevance literals. The language L^+ consists of domain rules, relevance rules, and integrity constraints. In L^+ there are no abducibles, and therefore its meaning is characterized in terms of stable models. Given a program T over L^+ and a literal L, we write $T \models L$ if L is true in every stable model of T. The program T is contradictory if $T \models \bot$.

Given a contradictory program T, to revise its contradiction ($\bot$) we modify T by adding and removing rules. In this framework, the diagnostic process reduces to finding such rules.

Given a set C of predicate symbols of L^+, C induces a partition of T into two disjoint parts: $T = T_c \cup T_s$. T_c is the changeable part and T_s the stable one. Let D be a pair $<U, I>$ where $U \cap I = \emptyset$, $U \subseteq C$ and $I \subseteq T_c$. We say that D is a diagnosis for T iff $(T - I) \cup U \models \bot$. $D = <U, I>$ is a minimal diagnosis if there exists no diagnosis $D_2 = <U_2, I_2>$ for T such that $(U_2 \cup I_2) \subset (U \cup I)$.

Example 7. Consider the following extension of the previous program Q, now expressed over L^+ with the introduction of integrity constraints:

$x \leftarrow a$
$x \leftarrow b$
$x \leftarrow c$
$expect(a)$
$expect(b)$
$expect(c)$

$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_1 \underline{v}$
$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_2 \underline{v}$
$a \lhd_1 b$
$b \lhd_1 c$
$b \lhd_2 a$
$\bot \leftarrow \underline{u} \lhd \underline{u}$
$\bot \leftarrow \underline{u} \lhd \underline{v}, \underline{v} \lhd \underline{u}$
$\bot \leftarrow \underline{u} \lhd \underline{v}, \underline{v} \lhd \underline{z}, not\ \underline{u} \lhd \underline{z}$

In this case, it holds that $T \models \bot$. Let $C = \{C_1, C_2\}$. T admits three minimal diagnoses:

$D_1 = <\{\ \}, \{a \lhd_1 b\}>$
$D_2 = <\{\ \}, \{b \lhd_1 c, b \lhd_2 a\}>$
$D_3 = <\{a \lhd_1 c\}, \{b \lhd_2 a\}>$

To compute the minimal diagnoses of a contradictory program T, we employ the contradiction removal method presented in [14], adapted here to handle relevancy relations. The contradiction removal method is based on the idea of revising (to false) some of the default atoms *not* A. A default atom *not* A can be revised to false by simply adding A to T. According to [14] the default literals *not* A that are allowed to change their truth value are those for which there exists no rule in T defining A. Such literals are called revisable.

Definition 9. *Let T be a program over L^+. An atom $A \neq \bot$ is a revisable of T iff there is no rule defining A in T.*

Definition 10. *Let T be a program over L^+ and V a set of revisables of T. A set $Z \subseteq V$ is a revision of T iff $T \cup Z \not\models \bot$.*

In order to apply the algorithm in [14], we need to transform our original program to one equivalent program which is suitable for contradiction removal.

Definition 11. *Let T be a program over L^+ and C a set of predicate symbols in L^+. The transformation Γ that maps T into a program T' is obtained by applying to T the following two operations:*

- *Add not incorrect($A \leftarrow Body$) to the body of each rule $A \leftarrow Body$ in T_c.*
- *Add the rule $p(x_1,...,x_n) \leftarrow uncovered(p(x_1,...,x_n))$ for each predicate symbol p with arity n in C, where $x_1,...,x_n$ are variables.*

We assume the predicate symbols *incorrect* and *uncovered* do not occur in T.[3]

Example 8. Let T be the program of Example 7. Then, the program $\Gamma(T)$ is:

$$x \leftarrow a$$
$$x \leftarrow b$$
$$x \leftarrow c$$
$$expect(a)$$
$$expect(b)$$
$$expect(c)$$

$$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_1 \underline{v}$$
$$\underline{u} \lhd \underline{v} \leftarrow \underline{u} \lhd_2 \underline{v}$$

$$\bot \leftarrow \underline{u} \lhd \underline{u}$$
$$\bot \leftarrow \underline{u} \lhd \underline{v}, \underline{v} \lhd \underline{u}$$
$$\bot \leftarrow \underline{u} \lhd \underline{v}, \underline{v} \lhd \underline{z}, not\ \underline{u} \lhd \underline{z}$$

$$a \lhd_1 b \leftarrow not\ incorrect(a \lhd_1 b)$$
$$b \lhd_1 c \leftarrow not\ incorrect(b \lhd_1 c)$$

3 The proof procedure for the correctness of the transformation Γ is detailed in [6].

$b \lhd_2 a \leftarrow not\ incorrect(b \lhd_2 a)$

$\underline{u} \lhd_1 \underline{v} \leftarrow uncovered(\underline{u} \lhd_1 \underline{v})$
$\underline{u} \lhd_2 \underline{v} \leftarrow uncovered(\underline{u} \lhd_2 \underline{v})$

$\Gamma(T)$ admits three minimal revisions with respect to the revisables of the form *incorrect*(.) and *uncovered*(.):

$Z_1 = \{incorrect(a \lhd_1 b)\}$
$Z_2 = \{incorrect(b \lhd_1 c), incorrect(b \lhd_2 a)\}$
$Z_3 = \{uncovered(\text{a} \lhd_1 c), incorrect(b \lhd_2 a)\}$

The following result relates the minimal diagnoses of a program T with the minimal revisions of $\Gamma(T)$.

Theorem 1. *Let T be a program over* L^+. *The pair* $D = \langle U, I \rangle$ *is a diagnosis for T iff*

$$Z = \{uncovered(A): A \in U\} \cup \{incorrect(A \leftarrow Body): A \leftarrow Body \in \text{I}\}$$

is a revision of $\Gamma(T)$, *where the revisables are literals of the form incorrect*(.) *and uncovered*(.). *Furthermore, D is a minimal diagnosis iff Z is a minimal revision.*

To compute the minimal diagnosis of a program T we consider the transformed program $\Gamma(T)$ and compute its minimal revisions. An algorithm for computing minimal revisions in such logic programs is given in [14].

10. Concluding Remarks

We have shown that preferences and priorities (they too a form of preferential expressiveness) can enact choices amongst rules and amongst abducibles, which are dependant on the specifics of situations, all in the context of theories and theory extensions expressible as logic programs with updates. These programs are executable by means of publicly available state-of-the-art systems, using available transformations provided here and elsewhere [14]. We have furthermore shown how preferences about knowledge extensions can be integrated with knowledge updates, and how they too fall under the purview of updating, again in the context of logic programs. Preferences about preferences are also adumbrated therein.

Although we have based our approach on the Stable Model semantics, we could just as easily have used the Well-Founded Semantics for a more skeptical preferential reasoning. Other logic program semantics are available too, such as the Revised Stable Model semantics, a two-valued semantics which resolves odd loops over default negation, arising from the unconstrained expression of preferences, by means of *reductio ad absurdum* [15]. Indeed, when there are odd loops over default negation in a program, Stable Model semantics does not afford the program with semantics.

Also, we need not necessarily insist on a strict partial order for preferences, but have indicated that different conditions may be provided. The possible alternative revisions, required to satisfy the conditions, impart a non-monotonic or defeasible reading of the preferences given initially. Such a generalization permits us to go beyond a simply foundational view of preferences, and allows us to admit a coherent view as well, inasmuch several alternative consistent stable models may obtain for our preferences, as a result of each revision.

The abductive process and the system of a priori preferences can be improved as well, in order to allow for the abduction of a set of abducibles. It will be necessary to specify new ways in which we can express preferences over these sets, but work is already well underway in this regard. Abduction of multiple literals is, however, already possible in the current system, by making single abducibles themselves stand for sets of "abducible" literals. One just needs to carefully model all the relevant combinations of literals that one would be inclined to expect.

Concerning model-based diagnosis, the integration of actions as possible abductions is also a must in order to tackle even more complex problems and applications, especially in diagnosis and prospective logic programming, as we must necessarily deal with pre- and post-conditions. Namely, the preconditions for an action to be abduced must be evaluated before the update for the action actually takes place, but the post-conditions must also be taken in consideration during the simulation and before the real action is executed.

Preferences over observables will also be desirable, since not every observation costs the same for the agent. Performing an X-Ray costs more than checking for tooth mobility or gingival pockets, and these differences should be modelled directly in the system. It would also be interesting to study more general ways of selecting the most interesting internal observations to pay attention to at each diagnosis step. Furthermore, abductive reasoning can be used to generate hypotheses of observations of events possibly occurring in the future along the lines of [1]. This ability will allow us to foresee the occurrence of new events which may or may not occur within some time interval, and thereby confirm or disprove an abduced explanation.

In [17], arguments are given as to how epistemic entrenchment can be explicitly expressed as preferential reasoning. And, moreover, how preferences can be employed to determine believe revisions, or, conversely, how belief contractions can lead to the explicit expression of preferences. [7] provides a stimulating survey of opportunities and problems in the use of preferences, reliant on AI techniques.

We advocate that the logic programming paradigm (LP) provides a well-defined, general, integrative, encompassing, and rigorous framework for systematically studying computation, be it syntax, semantics, procedures, or attending implementations, environments, tools, and standards. LP approaches problems, and provides solutions, at a sufficient level of abstraction so that they generalize from problem domain to problem domain. This is afforded by the nature of its very foundation in logic, both in substance and method, and constitutes one of its major assets.

Indeed, computational reasoning abilities such as assuming by default, abducing, revising beliefs, removing contradictions, preferring, updating, belief revision, learning, constraint handling, etc., by dint of their generality and abstract characterization, once developed can readily be adopted by, and integrated into, distinct topical application areas. No other computational paradigm affords us with the wherewithal for their coherent conceptual integration, all the while being the very vehicle that enables testing its specification, when not outright its very implementation [16]. Consequently, it merits

sustained attention from the community of researchers addressing the issues we have considered and outlined.

11. Acknowledgements

We thank Gregory Wheeler for extended discussion and comment on the philosophical material in this paper. We thank Joana Nogueira for preparing and supplying scientific information for the case study on differential medical diagnosis, and also a number of colleagues at DEIS, U. Bologna, for prior stimulating discussions.

References

1. M. Alberti, M. Gavanelli, E. Lamma, P. Mello, and P. Torroni. Abduction with Hypotheses Confirmation. Poster paper. *Proc. Intl. Joint Conf. on Artificial Intelligence (IJCAI-05)*, pages 1545-1546, 2005.
2. J. J. Alferes and L. M. Pereira. Updates plus preferences. In M. O. Aciego, I. P. de Guzmán, G. Brewka, and L. M. Pereira, editors, *Logics in AI, Procs. JELIA'00*, LNAI 1919, pages 345–360, Springer 2000..
3. J. J. Alferes, A. Brogi, J. A. Leite, L. M. Pereira. Evolving logic programs. In S. Flesca, S. Greco, N. Leone, G. Ianni, editors, *Procs. of the 8th European Conf. on Logics in Artificial Intelligence (JELIA'02)*, LNCS 2424, pages 50–61, Springer, 2002.
4. J. J. Alferes, L. M. Pereira, H. Przymusinska and T. C. Przymusinski. LUPS – a language for updating logic programs. *Artificial Intelligence*, 138(1–2), 2002.
5. G. Brewka and T. Eiter. Preferred answer sets for extended logic programs. *Artificial Intelligence*, 109:297–356, 1999.
6. P. Dell'Acqua and L. M. Pereira. Preferential theory revision. *Journal of Applied Logic*. Forthcoming, 2007.
7. Jon Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, 2004.
8. Pierre Duhem. *The Aim and Structure of Physical Theory*. Princeton University Press, 2nd edition, 1954.

9. M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In R. Kowalski and K. A. Bowen, editors, *ICLP'88*, pages 1070–1080. MIT Press, 1988.
10. Sven Hansson. Ten philosophical problems in belief revision. *Journal of Logic and Computation*, 13:37–49, 2003.
11. Isaac Levi. *Mild Contraction*. Clarendon Press, Oxford, 2004.
12. G. Lopes and L. M. Pereira. Prospective Logic Programming with ACORDA. In *Empirically Successful Computerized Reasoning (ESCoR'06)* workshop at the 3rd International Joint Conference on Automated Reasoning (IJCAR'06), August 2006.
13. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.
14. L. M. Pereira, C. Damásio and J. J. Alferes. Debugging by Diagnosing Assumptions. In P. Fritzson, editor, *1st Int. Ws. on Automatic Algorithmic Debugging, AADEBUG'93*, LNCS 749, pages 58–74, Springer, 1993.
15. L. M. Pereira and A. M. Pinto. Revised Stable Models - a Semantics for Logic Programs. *12th Portuguese Intl. Conf. on Artificial Intelligence (EPIA'05)*, LNAI 3808, pages 29–42, Springer, 2005.
16. L. M. Pereira, Philosophical Incidence of Logical Programming, in: Handbook of the Logic of Argument and Inference, D. Gabbay et al. (eds.), pages 425–448, Studies in Logic and Practical Reasoning series, volume 1, Elsevier Science 2002.
17. Hans Rott. *Change, Choice and Inference*. Oxford University Press, Oxford, 2001.
18. Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In William L. Harper and Brian Skyrms, editors, *Causation in Decision, Belief Change and Statistics*, volume 2, pages 105–134, Reidel, 1987.

Adaptive Logics for Abduction and the Explication of Explanation-Seeking Processes

Joke Meheus
Centre for Logic and Philosophy of Science
Ghent University
Joke.Meheus@UGent.be

1. Introduction

The aim of this paper is twofold. First, I want to illustrate the main characteristics of abductive reasoning processes by means of an example from the history of the sciences. Next, I want to show that such reasoning processes can be explicated by means of a formal logic.

The example is taken from the history of chemistry and concerns a very small episode from Lavoisier's struggle with the 'air' obtained from *mercurius calcinatus per se* (mercury oxide, in modern terminology). Eventually, this struggle would lead to the discovery of oxygen.

Readers who have a romantic view on scientific discovery and creativity should immediately be warned: they will probably be disappointed by the example. It is not a story about a series of genius insights that all of a sudden led to one of the most important discoveries in chemistry. To the contrary, the steps that I shall document on were all very small and quite mundane, and some may even question whether they were rational. However, what makes the example so fascinating is that it is one of those extremely rare occasions where one obtains some insight in the microstructure of reasoning—thanks to some notes that were preserved, it is possible to reconstruct, step by step, the individual inferences that Lavoisier made.

As the example will nicely show, abductive reasoning processes have two important characteristics. The first is that abductive steps are combined with

deductive steps. The second is that, partly because of this combination, abductive reasoning processes are *dynamic*. For instance, a conclusion reached on the basis of an abductive step may be withdrawn when its negation is derived by deductive means.

These characteristics confront logicians with the problem to design systems that combine (in a sensible way) deductive and abductive inference rules and that can moreover account for the dynamics involved. In Meheus et al. (2006), it is shown that so-called adaptive logics enable one to solve this problem. Adaptive logics are a specific kind of formal systems that are especially suited for the study of reasoning processes that are non-monotonic and/or dynamic.[1]

The logic presented in Meheus et al. (2006) is called $\mathbf{LA}^r$ and is an ampliative extension of first-order Classical Logic (henceforth **CL**). The logic is intended for the abduction of singular hypotheses and presupposes that, with respect to a specific application, the set of *explananda* and the set of possible *explanantia* are disjoint (but not necessarily exhaustive). Where an *explanandum* can be explained by different *explanantia*, $\mathbf{LA}^r$ allows only for the abduction of their disjunction.

In this paper, I shall show that the logic $\mathbf{LA}^r$ allows for the explication of Lavoisier's reasoning, and thus provides an insight in the logical structure of his reasoning. I shall also argue that this logical explication enables us to counter the claim that Lavoisier's reasoning was fallacious.

A second warning is in order. I am evidently not claiming that Lavoisier was using an adaptive logic. I am even not claiming that he had any explicit ideas about the logic behind his reasoning—he most likely had not. What I do believe, however, is that Lavoisier had sound logical intuitions, and that the logic $\mathbf{LA}^r$ provides us with a justification for this claim.

1 The first logic in this family was designed by Diderik Batens around 1980 and was meant to interpret (possibly) inconsistent theories as consistently as possible. Later the notion of an adaptive logic was generalized in different ways (for instance, to capture ampliative forms of reasoning) and a whole variety of adaptive logics was designed—for an overview, see Batens (2007).

2. Explaining the Properties of 'Oxygen'

Early March 1775,[2] Lavoisier conducted a series of experiments on the 'air' that is released when one reduces the red calx of mercury (mercury oxide) to mercury, without the addition of charcoal.[3] As Lavoisier believed that the red calx of mercury was mercury combined with *fixed air*,[4] he was convinced before beginning the experiments that the air he obtained from the reduction of the mercury calx would be fixed air—see Holmes (1985, p. 45). However, when he performed the standard limewater test for fixed air, he found out that the air obtained from the mercury calx did not form a precipitate (as it should have, had it been fixed air).

Lavoisier next turned to the other standard qualitative tests for 'airs'. Also these ruled out that the air was fixed air. As he wrote in his notebook,[5]

> One was first of all curious to test the effect of this air on animals. For this purpose one passed it into a jar into which one introduced a bird. One left the bird inside it for a good half minute, without its appearing to suffer there in the least. Removed from the air, it flew away without having suffered in any way. [Holmes (1985, p. 46)]

This test suggested that the air was very respirable. His experiments with burning candles further confirmed that the air involved was 'better' than common air:

> One repeated the experiment of the candle two times, and in large jars. It is charming. The flame is much larger and much clearer and much more

2 The historical context of the example is discussed at length in Holmes (1985); I refer to this work for more details.

3 As was known already by the Alchemists, liquid mercury can be converted into a red powder (called mercurius calcinatus per se) by heating it; by further heating this powder, the mercury can be recovered from it. The latter reduction can be done without the addition of charcoal which posed a problem for the phlogiston theory—see Holmes (1985). for more details. This was one of the reasons why Lavoisier, among others, was interested in the air that is released during the reduction from mercury oxide to mercury.

4 The term "fixed air" was introduced by Stephen Hales to refer to ordinary air in a 'fixed' state. The present-day term is carbonic acid gas.

5 This and the following notes were translated from French to English by Larry Holmes. I follow his translation.

> beautiful than in common air, but in color no different from an ordinary flame. [Holmes (1985, p. 47)]

Believing by now that the air was better than common air (in the sense that it better supported respiration and combustion than common air does), Lavoisier performed the *nitrous air test*.[6] This quantitative test was designed by Priestley and was considered by Lavoisier as more reliable than the qualitative tests. During the nitrous air test, one portion of nitrous air was mixed with two portions of the air one wanted to test. The diminution in volume provided an indication of the air one was dealing with. For instance, it was known that common air led to a diminution of one-fifth.

During the nitrous air test, there were actually two kinds of results. On the one hand, there were a number of qualitative results, such as the colour of the vapours that were formed and the rapidity of the effect. These results were available almost immediately after the experiment started. On the other hand, there were the quantitative results concerning the diminution of the volume of the mixed airs. These were available only at the end of the test.

The day that Lavoisier performed the nitrous air test (March 31, 1775), he entered the following note in his notebook:[7]

> One introduced [one part of] nitrous air into two parts of this air. It appeared that the red color of the vapors was more marked, and the effect more rapid than with common air.
>
> There was a diminution of

Lavoisier did not finish this sentence. This indicates that he wrote this note at the beginning of the experiment, when he did not yet know the amount of the diminution. He left a space open, where he could later fill in the amount of the diminution, and wrote below it

> so that, according to this operation, one could judge that this air is more perfect to common air.

6 In modern terminology, nitrous air is *nitrogen oxide*.

7 The note was discovered by Larry Holmes. I follow his analysis of the note as presented on p. 47 of Holmes (1985).

What happened here? The first qualitative results strengthened Lavoisier's belief that he was dealing with an air that is better than common air (the colour of the vapours was more intense and the effect more rapid that in the case of common air). This belief must have been so strong that he entered it as the conclusion of the entire experiment, even before he knew the quantitative results. From the way in which he wrote the note (leaving open a space to fill in later the outcome of the experiment), one may infer that Lavoisier was fully confident that the quantitative results would confirm the belief with which he started the experiment.

But then something fascinating happened. When he actually obtained the quantitative results, Lavoisier changed his mind about the conclusion. Instead of simply entering the amount of the diminution in the open space, he left the unfinished sentence unfinished, and entered instead all the numerical details (which previously he did not bother to mention):

> One employed two measures of this air,
> each 2.7 cubic inches, making together 5.4
> One added nitrous air 2.7
>
> 8.1 cubic inches
>
> The 8.1 cubic inches was reduced almost
> immediately to 4.42

Having thus filled the open space, he crossed out his first conclusion (namely that the air was more perfect than common air), and now wrote below it:

> That is to say, regarding the portion of nitrous air as probably entirely absorbed, there was one cubic inch, that is to say one-fifth, of the air absorbed.
>
> That is about the proportion of common air.

As is clear from this passage, Lavoisier gave priority to the quantitative results of the nitrous air test, and hence, concluded that the air obtained from mercury calx was simply common air.

Obviously, the story does not end here. The problem remained why the qualitative results indicated that the air was better than common air. Lavoisier

never arrived at a satisfactory solution to this problem. It was Priestley who eventually solved the problem by demonstrating, on the basis of a revised version of the nitrous air test, that the air obtained from mercury calx was indeed different from common air.

However, for this paper, the rest of the story is not important. In what follows, I shall concentrate on the elements that led to Lavoisier's note of March 31, 1775.

3. Analysis of Lavoisier's Reasoning

The qualitative tests that Lavoisier performed confronted him with two surprising facts concerning the sample of air obtained from the reduction of mercury calx:

F1 Some bird stayed in a sample of the air under investigation for more than half a minute without suffering.

F2 Some candles burned in a sample of the air under investigation with a larger flame than a normal candle does in a sample of common air.

These facts where surprising in view of Lavoisier's initial expectation that the air under investigation was fixed air (and hence that candles would extinguish in it and birds would die in it). Both facts, however, could easily be explained in view of the following generalizations:

G1 If some air is better than common air, then a bird can stay in it for more than half a minute without suffering.

G2 If some air is better than common air, then a candle burns in it with a larger flame than a normal candle does in common air.

These generalizations were common knowledge at that time and were accepted by Lavoisier. From **F1** and **R1**, one can derive, by abduction:

G2 The air under investigation is better than common air.

The same conclusion also follows abductively from **F2** and **G2**.

This is the conclusion that Lavoisier first entered in his notebook. This conclusion was further confirmed by the qualitative results of the nitrous air test. However, the nitrous air test also led to the following fact:

F3 A sample of the air under investigation led to a reduction of one-fifth in the nitrous air test.

As Lavoisier also accepted the following generalization:

G3 Some air is common air if and only if it leads to a reduction of one-fifth in the nitrous air test.

he was able to derive, *deductively*:

C2 The air under investigation is common air.

This is the conclusion which Lavoisier retained and for which he rejected the earlier conclusion **C1**.

4. Characterisics of Lavoisier's Reasoning Process

The reasoning process analysed in the previous section has several interesting properties. The first is that ampliative steps are combined with deductive steps in one and the same reasoning process. For instance, whereas **C1** is derived by means of abduction, **C2** is derived deductively.

The second property is that the reasoning process is *non-monotonic*: earlier conclusions are rejected in view of new information. As soon as **F3** is added as a new premise, Lavoisier rejects **C1**. This seems reasonable in view of the fact that **C1** was only derived *abductively* and that, from **F3** and **G3**, it follows, *deductively*, that **C1** does not hold true. Thus, ampliative conclusions are reviewed when a 'stronger' derivation to their negation is available.

A final characteristic is that inference rules are validated *contextually*. The abductive inference to **C1** is considered as valid *until* **C2** is derived. At that point, it is no longer considered as a valid inference. Lavoisier makes this clear by striking out its conclusion.

5. Logic-Based Approaches to Abduction

In the next section, I shall briefly present the adaptive logic **LA**^{r} and show that it is suitable for the explication of Lavoisier's reasoning. But first I argue why **LA**^{r} is better suited for this explication than the logic-based approaches to abduction which were developed in Artificial Intelligence.[8]

Within logic-based approaches, abductive inferences are perceived as falling under the following argumentation scheme:

†) $A \supset B$, B / A

This scheme, which is generally known as *Affirming the Consequent*, is evidently not deductively valid. Hence, as the framework of most logic-based approaches is a deductive one, the above scheme is not implemented directly. Instead, abductive inferences are specified as a kind of 'backward reasoning': given a theory T and an *explanandum* B, find an A such that

(1) $T \cup \{A\} \vdash B$.
(2) $T \nvdash B$
(3) $T \nvdash \neg A$.
(4) $B \nvdash A$.
(5) A is 'minimal'.

The first of these requirements needs no explanation. Also the next two requirements are straightforward: (2) warrants that the *explanandum* B is not explained by the background theory, and (3) that the explanatory hypothesis A is compatible with T.[9] (4) is needed to rule out degenerate cases. For instance, we do not want to abduce B as an explanation for itself. Also, if $T \cup \{A\} \vdash B$, then $T \cup \{A \vee B\} \vdash B$, but we do not want $A \vee B$ as an explanation for B. Cases like this are ruled out by requiring that the truth of the explanatory hypothesis is not warranted by the truth of the *explanan-*

8 In recent years, abductive reasoning has gained an enormous interest in the domain of Artificial Intelligence. At this moment, a large number of systems is available for a variety of application contexts: diagnostic reasoning, text understanding, case-based reasoning, planning , ... For an interesting overview of AI-approaches to abduction, see Gabriele (2000).

9 A formula A is said to be compatible with a set of premises Γ iff $\Gamma \nvdash \neg A$

dum—this is what (4) comes to. (5) is related to the fact that, when trying to explain some *explanandum*, one is interested in explanations that are as parsimonious as possible. Hence, in view of $A \supset B \vdash \mathbf{CL}(A \wedge D) \supset B$, one needs to prevent that $A \wedge D$ can be abduced, whenever A can. This can be realized by requiring that the explanatory hypothesis is 'minimal'. This notion of minimality can be defined in different ways—one may, for instance, consider an explanatory hypothesis as minimal if no alternative is available that is logically weaker. However, no matter how it is defined, minimality is a *comparative* notion: whether some explanatory hypothesis A is minimal with respect to some *explanandum* B depends on the available alternatives.

It is important to note that several of the above requirements are *negative*. This does not only hold true for (2)–(4), but also for (5). Indeed, as minimality is a comparative notion, (5) entails:

(5') $T \cup \{C\} \nvdash_{\mathbf{CL}} B$, for every C that satisfies (2)–(4) and in view of which B is not minimal.

One consequence of these negative clauses was already mentioned in the previous section: the consequence relation defined by (1)–(5) is *non-monotonic*. Conclusions that follow abductively from some theory T may be withdrawn when T is extended to $T \cup T'$.

Another consequence is that, at the predicative level, the consequence relation defined by (1)-(5) is not only undecidable, there even is no positive test for it[10]. This is related to the fact that first-order predicate logic is undecidable—if some conclusion A does *not* follow from a set of premises Γ, we may not be able to establish this. Hence, as the consequence relation is partly defined in terms of negative requirements, it immediately follows that, for undecidable fragments, it lacks a positive test. Suppose, for instance, that for some theory T, some *explanandum* B and some sentence A, (1) is satisfied. In that case, it seems reasonable to conclude that A follows abductively from T. However, if one is unable to establish that also (2)–(5) are satisfied, no reasoning can warrant that this conclusion is not erroneous.

10 Even if A follows abductively from a theory T and an *explanandum* B, there need not exist any finite construction that establishes this.

There are different ways to deal with the lack of a positive test. The one usually followed within Artificial Intelligence is to consider only decidable fragments of first-order logic. The rationale behind this is clear: when dealing with decidable fragments, one may be sure that, for arbitrary theories T and *explananda* B, there is an algorithm for (2)–(5), and hence, that a decision method can be designed for "follows abductively from". From the point of view of applications, however, this is an enormous restriction: many interesting theories are undecidable.

An alternative way is to allow that inferences are made, not on the basis of absolute warrants, but on the basis of one's best insights in the premises. When this second option is followed, abductive reasoning processes not only exhibit an *external* form of dynamics (adding new information may lead to the withdrawal of previously derived conclusions), but also an *internal* one (the withdrawal may be caused by merely analysing the premises). Suppose, for instance, that for some theory T, some *explanandum* B, and some sentence A, one established that (1) is satisfied, and one did not establish that one of (2)–(5) is violated. In that case, it seems rational to conclude that A follows abductively from T. This conclusion, however, is provisional. If at a later moment in time, one is able to show that one of the negative requirements is violated (for instance, because one established that $\neg A$ follows from T), A has to be withdrawn as an explanation for B.

There are several arguments in favour of this second option. The first is that unwanted restrictions are avoided: abduction can be defined for *any* first-order theory. A second argument is that the conclusions of abductive reasoning processes are defeasible anyway. Whether the withdrawal of a conclusion is caused by an external factor or an internal one does not seem to be essential. The third, and most important argument is that, even for decidable fragments, it is often unrealistic to require absolute warrants. Even if a decision method is available, reasoners may lack the resources to perform an exhaustive search, and hence, may be forced to act on their present best insights.

The logic $\mathbf{LA}^r$ follows the second option. This has the advantage that, even for undecidable fragments, it enables one to come to justified conclu-

sions. These conclusions are tentative and may later be rejected, but they constitute, given one's insight in the premises at that moment, the best possible estimate of the conclusions that are 'finally derivable' from the premises.[11]

The logic **LA**^{r} has several other advantages. A first one is that (unlike the systems developed within Artificial Intelligence) it has a *proof theory*. As we shall see below, this proof theory is dynamic (conclusions derived at some stage may be rejected at a later stage), but it warrants that the conclusions derived at a given stage are justified in view of the insight in the premises at that stage. This is especially important as, at the predicative level, there is no positive test for abductive reasoning.

Another advantage of the proposed logic is that it is much closer to natural reasoning than the existing systems. As was mentioned in the beginning of this section, abduction is usually viewed as a form of backward reasoning—"find an A that satisfies the requirements (1)–(5)". The search procedure by which this is realized in the existing systems (for instance, some form of linear resolution) is very different from the search procedures of human reasoners. The logic **LA**^{r} treats abduction as a form of 'forward reasoning': it is an ampliative system that directly validates inferences of the form **(†)**.

The third advantage is related to this: unlike what is the case for the AI approaches to abduction, deductive and abductive steps are nicely integrated into a single system. As a consequence, the logic not only enables one to generate explanatory hypotheses, but also to infer predictions on the basis of explanatory hypotheses and the background theory. This is highly important from the point of view of applications. In scientific contexts, for instance, explanatory hypotheses are typically used to derive predictions which, in turn, may lead to a revision of the original hypotheses.

11 Roughly speaking, an 'abductive conclusion' A is finally derivable from a theory T if the requirements (1)-(5) are satisfied—see the next section for a precise definition of this notion.

6. An Adaptive Logic for Abduction

The general idea is extremely simple: it is allowed that the predicative version of (†), namely

(‡) $\quad A(\beta)$, $(\forall\alpha)(A(\alpha) \supset B(\alpha))$ / $B(\beta)$

is applied "as much as possible". For the moment, this ambiguous phrase may be interpreted as "unless and until $(\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))$ turns out to be **CL**-derivable from Γ". So, whenever it is **CL**-derivable from Γ that, for some general rule $(\forall\alpha)(A(\alpha) \supset B(\alpha))$ and some explanandum $B(\beta)$, (‡) cannot be applied consistently (because, $\neg A(\beta)$ is **CL**-derivable from Γ), the application of (‡) is overruled. In view of what we have seen in the previous sections, this is exactly what we want.

There is one general restriction, which is needed to obtain a sensible system. Where $\mathcal{W}$ is the set of closed formulas of the standard predicative language, one needs two sets of truth functions of closed primitive formulas,[12] $\mathcal{W}^e$ and $\mathcal{W}^a$, such that no primitive formula occurs in a member of $\mathcal{W}^e$ as well as in a member of $\mathcal{W}^a$. The sets may but need not be combinatorially closed, in other words, they need not contain all subformulas of their members or all truth-functions of these subformulas.

Intuitively, $\mathcal{W}^e$ is the set of *explananda*, formulas that are considered as requiring an explanation, whereas $\mathcal{W}^a$ is the set of *explanantia*, formulas that, if they can be abduced, form potential explanations for the *explananda*. The requirement that no primitive formula occurs in members of both sets can be easily justified with respect to applications. If one tries to abduce an explanation, one has in mind a phenomenon for which an explanation is sought, and the explanation should be logically independent of the explained phenomenon—everyone rejects (even partial) self-explanations. To save space, expressions of the form $(\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))$ will be abbreviated as $[\![B(\beta), \neg A(\beta)]\!]$ and, in line with what is common for adaptive logics, the formula $[\![B(\beta), \neg A(\beta)]\!]$ will be called an "abnormal-

12 Primitive formulas are those that contain no logical symbols, except possibly for identity.

ity".[13] As we will see below, it is possible that a disjunction of abnormalities is **CL**-derivable from a set of premises Γ without any of its disjuncts being derivable from it.

LA^r can be formulated in the standard format from Batens (2007), which greatly simplifies the technical matters. An adaptive logic **AL** is in standard format if it is characterized as a triple consisting of three elements: (i) **LLL**, a compact and monotonic lower limit logic, (ii) Ω, a set of abnormalities that all have the same logical form, and (iii) an adaptive strategy.

The lower limit logic **LLL** determines the part of the adaptive logic **AL** hat is not subject to adaptation. From a proof theoretic point of view, the lower limit logic delineates the rules of inference that hold unexceptionally. From a semantic point of view, the adaptive models of a premise set Γ are a selection of the **LLL**-models of Γ. The lower limit logic of **LA**^{r} is **CL**, and remember that its premise set is $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle$.

Abnormalities are formulas that are presupposed to be false, unless and until proven otherwise. Ω comprises all formulas of a certain (possibly restricted) logical form. In the case of **LA**^{r} the restriction will refer to $\mathcal{W}^e$ and $\mathcal{W}^a$. and the set of abnormalities Ω is defined as $\{(\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge (\neg A(\beta)) \mid A(\beta) \in \mathcal{W}^a; B(\beta) \in \mathcal{W}^e; \nvdash_{\mathbf{CL}} (\forall\alpha)(A(\alpha) \supset B(\alpha))\}$. In the present extensional framework, $(\forall\alpha)(A(\alpha) \supset B(\alpha))$ can be taken to express that A contains a (sufficient) cause for B—I write "contains" because A may itself be a conjunction and some of its conjuncts may not be required for warranting B. The second conjunct of an abnormality states that the specific sufficient cause $A(\beta)$ for $B(\beta)$ did not occur. The requirement that the generalization $(\forall\alpha)(A(\alpha) \supset B(\alpha))$ is not a **CL**-theorem has to be added in order to prevent that all models would display abnormalities. However, as this rules out at once cases in which $A(\beta)$ is a contradiction or $B(\beta)$ is a tautology, this requirement is harmless. (Nobody wants to seek an explanation for a tautology and nobody will accept an explanation by *Ex Falso Quodlibet*.) An adaptive logic presupposes that abnormalities are false unless and until proven otherwise. So, the presupposition of **LA**^{r} is that, if

13 The term "abnormality" refers to formulas that overrule the application of some desired inference rule—in our case the abduction scheme (‡).

an effect did occur, then all its potential causes (in the weak, extensional, sense) did also occur.

The strategy is Reliability. This strategy warrants that, in cases where more than one explanatory hypothesis can be abduced for the same *explanandum*, only their disjunction is derivable by $\mathbf{LA}^r$. It also warrants that, in cases where there are mutually inconsistent explanatory hypotheses, only those explanations are abduced that are *jointly compatible* with the premises. Both cases will be illustrated in the example at the end of this section.[14]

If one adds to the lower limit logic an axiom schema excluding that abnormalities occur, viz. an axiom schema that reduces abnormal premise sets to triviality, one obtains the so-called upper limit logic. The upper limit logic of $\mathbf{LA}^r$ is somewhat unusual as it refers to the sets $\mathcal{W}^e$ and $\mathcal{W}^a$. It is obtained by extending **CL** with the axiom schema $(\forall\alpha)(A(\alpha)\supset B(\alpha))\supset(B(\beta)\supset A(\beta))$ *provided* $B(\beta)\in\mathcal{W}^e$ and $A(\beta)\in\mathcal{W}^a$. It is easily seen that this comes to the requirement that, if the proviso is met, $(\forall\alpha)(A(\alpha)\supset B(\alpha))$ is logically equivalent to $(\forall\alpha)(A(\alpha)\equiv B(\alpha))$. As the upper limit logic is not interesting in itself, I shall not bother to give it a name.

Let us now turn to the proofs. If the deduction rules are formulated in generic format, they are identical for all adaptive logics in standard format. Let Γ contain the (declarative) premises as before, let the notation

$$A \quad \Delta$$

abbreviate that A occurs in the proof on the condition Δ, and let $Dab(\Delta)$ be the disjunction of the members of a finite $\Delta\subset\Omega$. The rules may be phrased as follows:[15]

PREM If $A\in\Gamma$

$$\frac{\ldots \quad \ldots}{A \quad \varnothing}$$

14 As is illustrated in Meheus et al. (2006); the Reliability Strategy also guarantees that, if the antecedent of some generalization has been arbitrarily strengthened, only sensible explanations are abduced.

15 The only rule that introduces non-empty conditions is RC. In other words, before RC is applied in a proof, the condition of every line will be $\varnothing$.

RU If $A_1,\ldots,A_n \vdash_{\mathrm{CL}} B$

$$\begin{array}{ll} A_1 & \Delta_1 \\ \ldots & \ldots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \end{array}$$

RC If $A_1,\ldots,A_n \vdash_{\mathrm{CL}} B \vee Dab(\Theta)$

$$\begin{array}{ll} A_1 & \Delta_1 \\ \ldots & \ldots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \cup \Theta \end{array}$$

In addition to the inference rules, also a *marking definition* is needed. The marking definition determines which lines of a proof have to be marked. Formulas that occur on marked lines are no longer considered to be derived in the proof.

I shall say that $Dab(\Delta)$ is a *minimal* Dab*-formula* at stage s of a proof if, at that stage, $Dab(\Delta)$ occurs in the proof on the empty condition and, for any $\Delta' \subset \Delta$, $Dab(\Delta')$ does not occur in the proof on the empty condition. Where $Dab(\Delta_1)$, ..., $Dab(\Delta_n)$ are the minimal Dab-formulas at stage s of the proof, $U_s(\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle) = \Delta_1 \cup \ldots \cup \Delta_n$ is the set of unreliable formulas at stage s. The marking definition for the Reliability Strategy is as follows:

Definition 1 *Line* i *is marked at stage* s *iff, where* Δ *is its condition,* $\Delta \cap U_s(\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle) \neq \varnothing$.

If $Dab(\Delta)$ is a minimal Dab-formula at stage s of the proof, then, in as far as one knows in view of the proof at this stage, the premises require one of the abnormalities in Δ to be true but do not specify which one is true. The Reliability Strategy considers all of them as unreliable. So the underlying idea is: if the understanding of the premises provided by the present stage of the proof is correct, the formulas occurring at unmarked lines are derivable from the premises, whereas the formulas occurring at marked lines are not.

Apart from the unstable derivability at a stage, one wants a stable kind of derivability, which is called final derivability.

Definition 2 *A is finally derived from $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle$ on line i of a proof at stage s iff (i) A is the second element of line i, (ii) line i is not marked at stage s, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.*

Definition 3 $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle \vdash_{\mathbf{LA}^r} A$ *(A is finally $\mathbf{LA}^r$-derivable from Γ) iff A is finally derived on a line of a $\mathbf{LA}^r$-proof from $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle$.*

Remark that these are definitions, and that they are not intended to have a direct computational use.

For the semantics of $\mathbf{LA}^r$, I refer the reader to Meheus et al. (2006). There, it is also shown that the semantics is sound and complete with respect to the dynamic proof theory.

The rest of this section is devoted to an illustration of the proof theory. I shall present a very simple example and not bother too much about technicalities. I shall concentrate on showing (i) that the logic leads to a nice integration of deductive and abductive steps, (ii) that it can handle the dynamics that is typical of abductive reasoning processes, and (iii) that the inference rule which corresponds to abduction is validated contextually.

Suppose that our set of premises Γ consists of the following generalizations

$$(\forall x)(Px \supset Rx)\ ,\quad (\forall x)(Px \supset Sx)\ ,\quad (\forall x)(Qx \supset Sx)\ ,\quad (\forall x)(Qx \supset Tx)\ ,\quad (\forall x)(\neg Px \supset Tx)$$

and the following data

$$Ra\ ,\ Rb\ ,\ \neg Sb\ ,\ Sc\ ,\ Sd\ ,\ \neg Td\ ,\ Re\ ,\ Te$$

Let $\mathcal{W}^e$ be the set of all singular formulas that are truth-functions of primitive formulas containing the predicates R, S and T, and $\mathcal{W}^a$ the set of all singular formulas that do not contain these predicates.

One way to start a $\mathbf{LA}^r$-proof from Γ is by entering all the premises:

1	$(\forall x)(Px \supset Rx)$	PREM	$\varnothing$
2	$(\forall x)(Px \supset Sx)$	PREM	$\varnothing$
3	$(\forall x)(Qx \supset Sx)$	PREM	$\varnothing$
4	$(\forall x)(Qx \supset Tx)$	PREM	$\varnothing$
5	$(\forall x)(\neg Px \supset Tx)$	PREM	$\varnothing$
6	Ra	PREM	$\varnothing$
7	Rb	PREM	$\varnothing$
8	$\neg Sb$	PREM	$\varnothing$
9	Sc	PREM	$\varnothing$
10	Sd	PREM	$\varnothing$
11	$\neg Td$	PREM	$\varnothing$
12	Re	PREM	$\varnothing$
13	Te	PREM	$\varnothing$

For each of these lines, the third element forms the "justification" for the formula that constitutes the second element. It contains the line numbers of the formulas from which the formula is derived (obviously empty in the case of premises) as well as the name of the rule by means of which the formula is derived (in the above case the premise rule PREM). The sets at the end of each line are the conditions—also these are obviously empty in the case of premises.

We are now in a position to make inferences from the premises. Let us first concentrate on the explanandum Ra. As is easily observed, the first generalization can be used to 'abduce' an explanatory hypothesis for Ra. In an $\mathbf{LA}^r$-proof from Γ, this is done by applying the rule RC:

14	Pa	$1,6$; RC	$\{[\![Ra, \neg Pa]\!]\}$

RC allows one to add abductive hypotheses to the proof, but only on a certain condition. This condition is represented by the fifth element of the line. Intuitively, line 14 can be read as: Pa is derivable from the formulas on lines 1 and 6, *unless and until* it can no longer be assumed (consistently) that $[\![Ra, \neg Pa]\!]$ is *false*.

Given our present insights in the premises (represented by the formulas that are explicitly written down in the proof), there is no reason to believe that $\neg Pa$ is true, and hence, it is consistent to assume that $[\![Ra, \neg Pa]\!]$ is false. This is why, at this stage of the proof, Pa is considered to be derivable from the premises (in view of line 14). If, at a later stage of the proof, it would turn out that the condition of line 14 is no longer satisfied, then this line will be 'marked' and the formula that occurs on it will no longer be considered to be derived. (The marking of lines will be illustrated below.)

In view of the formula on line 14, the second generalization allows one to infer the prediction Sa; this is done by means of the rule RU:

15 Sa $2,14$; RU $\{[\![Ra, \neg Pa]\!]\}$

RU is a generic rule that allows one to infer all **CL**-consequences: whenever some formula A is **CL**-derivable from a number of formulas $B_1, \ldots, B_n$ that are considered to be derived in the proof at some stage, then, at that stage, A can be added to the proof by means of RU. Note that RU is an *unconditional* rule: unlike RC, it does not lead to the introduction of new conditions. If, however, some of the B_i to which RU is applied are themselves derived on a non-empty condition, then these conditions are conjoined for the conclusion. Thus, as the formula of line 14 is used to derive the formula on line 15, the condition of the former is 'carried over' to the latter. This is obviously as it should be: if, at a later stage in the proof, the conclusion of line 14 is withdrawn because its condition is no longer satisfied, then all formulas that rely on it should also be withdrawn.

This is a first illustration of the way in which abductive steps and deductive steps are integrated. The rule RC allows one to generate new explanatory hypotheses (for instance, the one on line 14), and RU allows one to derive predictions from these.

Let us now turn to the explanandum Rb. As in the previous case, the rule RC enables us to abduce an explanatory hypothesis for Rb (see line 16 below). However, this time, we are also able to infer, by means of RU, the *negation* of our explanatory hypothesis:

16 Pb $1,7$; RC $\{[\![Rb, \neg Pb]\!]\}$
17 $\neg Pb$ $2,8$; RU $\varnothing$

Hence, we are able to infer the following abnormality:

18 $[\![Rb, \neg Pb]\!]$ $1,7,17$; RU $\varnothing$

At this stage in the proof, the condition of line 16 is no longer satisfied. As a consequence, the conclusion of line 16 is *withdrawn* from the proof. The withdrawal of a conclusion from the proof is recorded by *marking* the line on which the formula occurs. This is how the proof looks like at stage 18 (lines 1 to 15 are as before):

…

16 Pb $1,7$; RC $\{[\![Rb, \neg Pb]\!]\}$ $\surd^{18}$

17 $\neg Pb$ $2,8$; RU $\varnothing$

18 $[\![Rb, \neg Pb]\!]$ $1,7,17$; RU $\varnothing$

I shall now show what happens when more than one explanatory hypothesis can be abduced for the same explanandum. Have a look at Sc. In view of the relevant generalizations, the proof can be extended as follows:

19 Pc $2,9$; RC $\{[\![Sc, \neg Pc]\!]\}$

20 Qc $3,9$; RC $\{[\![Sc, \neg Qc]\!]\}$

However, as the reader can verify, the following disjunctions of abnormalities are **CL**-derivable from the premises:

21 $[\![Sc, \neg Pc \vee Sc, \neg(Qc \wedge \neg Pc)]\!]$ $2,3,9$; RU $\varnothing$

22 $[\![Sc, \neg Qc \vee Sc, \neg(Pc \wedge \neg Qc)]\!]$ $2,3,9$; RU $\varnothing$

The formula on line 21 expresses that $[\![Sc, \neg Pc]\!]$ or $[\![Sc, \neg(Qc \wedge \neg Pc)]\!]$ is true. Hence, it cannot be assumed that both disjuncts are false.

In view of such a disjunction of abnormalities, different strategies are possible. The one followed by $\mathbf{LA}^r$ is very cautious. As (at this stage of the proof) it is unclear which one of the two disjuncts is true, both disjuncts are (at this stage of the proof) considered as 'unreliable'. As a result, all formulas that are derived on the assumption that one of these disjuncts is false, are withdrawn. Thus, in our case, the formula on line 19 is withdrawn in view of the formula on line 21. By an analogous reasoning, the formula on line 20 is withdrawn in view of the formula on line 22:

…			
19	Pc	$2,9$; RC	$\{[\![Sc, \neg Pc]\!]\}$ $\surd^{21}$
20	Qc	$3,9$; RC	$\{[\![Sc, \neg Qc]\!]\}$ $\surd^{22}$
21	$[\![Sc, \neg Pc \vee Sc, \neg(Qc \wedge \neg Pc)]\!]$	$2,3,9$; RU	$\varnothing$
22	$[\![Sc, \neg Qc \vee Sc, \neg(Pc \wedge \neg Qc)]\!]$	$2,3,9$; RU	$\varnothing$

A mark may be removed at a later stage. Suppose, for example, that $[\![Sc, \neg(Qc \wedge \neg Pc)]\!]$ is **CL**-derivable from the premises, and is actually derived in the proof. So it would be clear which of the two disjuncts of the formula of line 21 is true, viz. the second one. As a result, line 19 would not be marked any more (unless $[\![Sc, \neg Pc]\!]$ is a disjunct of another disjunction of abnormalities).

As we have seen, apart from derivability at a stage, a stable notion of derivability is defined, viz. final derivability. Intuitively, a formula is finally derived on line i of a proof iff it is possible to extend the proof in such a way that line i is unmarked and remains unmarked in every further extension of the proof.

In view of the present premises, lines 19 and 20 will remain marked in any extension of the proof. So neither Pc nor Qc is finally derivable from the premises. However, their disjunction $Pc \vee Qc$ is. This can be seen from the following extension of the proof:

23	$(\forall x)((Px \vee Qx) \supset Sx)$	$2,3$; RU	$\varnothing$
24	$Pc \vee Qc$	$9,23$; RC	$\{[\![Sc, \neg(Pc \vee Qc)]\!]\}$

As no minimal disjunction of abnormalities is derivable that has $[\![Sc, \neg(Pc \vee Qc)]\!]$ as one of its disjuncts, the formula on line 24 is finally derivable from the premises.

Also for the explanandum Sd the rule RC enables one to derive a disjunction of explanatory hypotheses:

25	$Pd \vee Qd$	$2,3,10$; RC	$\{[\![Sd, \neg(Pd \vee Qd)]\!]\}$

This time, however, one of the disjuncts can be eliminated by pure deductive means:

26	$\neg Qd$	$4,11$; RU	$\varnothing$
27	Pd	$25,26$; RU	$\{[\![Sd, \neg(Pd \vee Qd)]\!]\}$

This again nicely illustrates how $\mathbf{LA}^r$ allows for the integration of deductive and abductive steps.

Let us finally turn to the situation where different explanatory hypotheses are mutually incompatible with the premises. As may be seen from the following extension of the proof, this is the case for the explanatory hypotheses that are abducible for Re and Te:

28	Pe	$1,12$; RC	$\{[\![Re, \neg Pe]\!]\}$
29	$\neg Pe$	$5,13$; RC	$\{[\![Te, Pe]\!]\}$

Although both these hypotheses may be entered at some stage in the proof, neither of them is finally derivable from the premises. This is warranted by the following $\mathbf{CL}$-derivable disjunction of abnormalities:

30	$[\![Re, \neg Pe]\!] \vee [\![Te, Pe]\!]$	$1,5,12,13$; RU	$\varnothing$

As soon as the formula on line 30 is added to the proof, lines 28 and 29 are marked—they remain marked in any extension of the proof.

7. Was Lavoisier's Reasoning Rational?

To some it may seem that Lavoisier, in the course of that particular experiment in March 1775, was not reasoning in a rational way. As some may argue, Lavoisier not only committed the fallacy of "affirming the consequent", he moreover jumped to the conclusion on the basis of partial information.

In the absence of formal logics for abduction, this argument carries some weight. It seems good practice to link rationality to "reasoning according to the standards of an appropriate logic". Hence, if there are doubts about the logicality of a particular inference form, this immediately casts doubts on its rationality.

In the case of abduction, the suspicion seemed justified. Not only is the inference not deductively valid, many examples of purportedly sound abduc-

tions seem to rely on a hidden non-formal reasoning. Indeed, the only sensible formal rule behind them seems to lead inevitably to a set of unsound and even inconsistent conclusions. For instance, from the explananda Qa and Ra and the generalizations $(\forall x)(Px \supset Qx)$ and $(\forall x)(\neg Px \supset Rx)$, ($\ddagger$) enables one to generate both Pa and $\neg Pa$.

In the previous section, we have seen, however, that it is possible to design a formal logic for abduction that is just as rigourous as, say, Classical Logic. What is important about this logic is that it enables one to distinguish between sound and unsound applications of ($\ddagger$). We have also seen that this distinction is *contextual*: even if abduction should be invalidated with respect to some of the premises, it may be validated with respect to others. Thus, where the generalizations are $(\forall x)(Px \supset Qx)$, $(\forall x)(\neg Px \supset Rx)$ and $(\forall x)(Sx \supset Rx)$, and the *explananda* are Qa and Ra, abduction should be invalidated with respect to the first two premises (neither Pa nor $\neg Pa$ should be abducible), but should be validated with respect to the third premise (Sa should be abducible). This is precisely what the logic **LA**^{r} allows for.

Applied to Lavoisier's premises, the logic **LA**^{r} allows one to derive **C1** from **F1** and **G1** as well as from **F2** and **G2**. Moreover, as long as **F3** is not added, the conclusion **C1** is *finally derivable* from the premise set. It is only when the fact **F3** is added (together with the generalization **G3**) that the inference to **C1** is invalidated. (In an **LA**^{r}-proof, this would be expressed by marking the line on which **C1** is derived.)

Some may still have problems with the fact that Lavoisier derived the conclusion *before* he knew all the results to which the experiment would lead. We have seen, however, that the inference to **C1** was not a deductive one, and that it was not treated by Lavoisier as such. He remembered very well that the inference was provisional and he (literally) deleted it as soon as its negation was derived deductively. This is again as one would expect on the basis of **LA**^{r}. Moreover, as I argued in Section 4, making provisional inferences is the only sensible way to proceed in cases where there is no positive test.

What is important is that one remembers the conditions under which a provisional judgment should be revised. Abductive inferences should be considered as fallacious if, and only if, these conditions are not remembered.

Acknowledgements

Research for this paper was supported by subventions from Ghent University and from the Research Foundation—Flanders (FWO - Vlaanderen).

References

Diderik Batens (2007). A universal logic approach to adaptive logics. *Logica Universalis*, 1:221--242.

Paul Gabriele (2000). AI approaches to abduction. In Dov~M. Gabbay and Philippe Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Volume 4: Abductive Reasoning and Learning*, pages 35--98. Kluwer Academic Publishers, Dordrecht.

Frederic Lawrence Holmes (1985). *Lavoisier and the Chemistry of Life. An Exploration of Scientific Creativity*. The University of Wisconsin Press.

Joke Meheus and Diderik Batens (2006). A formal logic for abductive reasoning. *Logic Journal of the IGPL*, 14:221-236.

Content Guidance in Formal Problem Solving Processes.

Diderik Batens
Centre for Logic and Philosophy of Science
Ghent University
Diderik.Batens@rug.ac.be

1. Some Background

On the present-day view of scientific rationality, science is content-guided. As Dudley Shapere phrases it in [62]: scientific inquiry relies on "what we have learned, including what we have learnt about how to learn" (p.52). While the Vienna Circle ultimately reduced scientific methodology to an a priori matter, aiming at delineating all possible science, the present view understands science as essentially relying on contemporary insights in the world as well as in methodological matters. While Kuhn and other members of the 'historicist' movement reduced science to a form of relativism, the present view sees scientific decisions as justified in terms of those contemporary insights.

It may come as a surprise that this view requires a heavy import of logic. In trying to understand scientific reasoning, especially its more creative aspects, heuristic aspects play a central role. One should clearly distinguish between heuristic aspects that pertain to logic and other heuristic aspects. Both kinds are important (and both result from a learning process). Even in order to give the non-logical aspects their due place, the logical aspects should be located and systematically described.

So I shall describe an approach to problem solving that has a remarkable combination of properties. On the one hand, the approach is formal in a specific but plain sense of the term. Yet, at the same time, it leaves ample

room for content-guidance. Central to the approach are formal problem solving processes—henceforth *fpsps*. These are defined by ingredients of different sorts: one or more logics (for declarative statements), the prospective dynamics of these logics, one or more erotetic logics, an oracle, a procedure (set of instructions), and a set of heuristic instructions. The main aim is obviously the *explication* of actual problem solving processes. The backbone is to solve a problem of the form $\{?\{A, \sim A\}\}$ by deriving one of its direct answers, A or $\sim A$, from Γ by Classical Logic—henceforth **CL**.

The logics for declarative statements should not be seen as alternatives for **CL** (Classical Logic). They are formal explications of methodological reasoning forms. These are handled by adaptive logics, of which there are two sorts: corrective and ampliative. Corrective adaptive logics are invoked for handling premise sets that display unexpected and unwanted properties, such as inconsistency and ambiguity. In the presence of these, **CL** results in triviality, while corrective adaptive logics interpret the premise set as consistently as possible or as unambiguously as possible. Ampliative adaptive logics extend **CL** in view of methodologically justifiable forms of reasoning: inductive generalization, abduction (see Meheus' paper in the present volume), compatibility reasoning (for example for extending a theory), etc.[1]

Most adaptive logics require defeasible inferences. Indeed, the reasoning forms have the peculiarity that there is no positive test for them. This makes their proofs necessarily dynamic: it may be required that a step derived at some point is considered as OUT at a later point and possibly as IN again at a still later point. The reasons for considering steps as IN or OUT are provided by the insights from the ongoing reasoning. Such reasoning forms are explicated by the dynamic proofs of adaptive logics—see [12,16,19] for an introduction as well as for technical results, and Section 3 for a brief introduction. In those dynamic proofs, the dynamics is technically controlled by conditions and marking definitions.

1 Whether an adaptive logic is corrective or ampliative is defined, in the present context, with respect to **CL**, which is here taken as the standard of deduction. This is merely a pragmatic convention.

A further dynamics is introduced. The prospective dynamics enables one to push formal elements of the proof search into the proof itself. This allows for more or less permissive heuristic procedures and makes the heuristic aspects transparent. The prospective dynamics can be spelled out for the dynamic proofs of adaptive logics.

Erotetic logics (in the style of [66]) are used to write problems (sets of questions) into the proof and to derive problems from given problems and declarative (prospective) statements. The erotetic logics handle the introduction of subproblems and so-called derived problems.

The framework is easily extended (i) with empirical means, observation and experiment, and (ii) with the introduction of available information, which originally was not seen as relevant. Problems that can be solved by empirical means are handled by a specific kind of oracle. Some problems will only be answered in the context of an experimental setup, which may require a fpsp itself.

Interestingly, both extensions are guided by the fpsps: the present state of the fpsp leads to an appeal to empirical means or to the introduction of new information. This guidance has two sources. On the one hand, the prospective dynamics may reveal (in terms of the positive part relation from Section 2) that available information is relevant for the problem. The information may be knowledge that was previously not seen as relevant, or it may be obtained by known empirical means. On the other hand, the derivable disjunctions of abnormalities (jargon to be explained below), which play an essential role in adaptive logics, sensibly guide research in suggesting empirical questions as well as defeasible theoretical choices. Both lead to useful steps towards the solution of the problem.

All these elements are joined and governed by a procedure. The specific road followed to solve a problem is determined by choices that are permitted by the procedure. These choices make it possible to proceed in view of "what we have learnt about how to learn." I shall pay special attention to further aspects that make fpsps content-guided.

In order to keep the discussion within bounds, I shall only consider the technicalities for the propositional case. This is merely a pragmatic decision.

The predicative version of the adaptive logics is available, and the same holds for the prospective dynamics.

The background of the approach is formed by many sources of inspiration. Science is seen a problem-solving activity, a widespread view that apparently reached a stable formulation in [36]. From philosophy of science and epistemology, this view is elaborated by a combination of ideas mainly taken from [38,39,40,41,42,43], which combine Tom Nickles' view on discovery in terms of constrained problem solving processes, [51,52,53,54,55,56,57] with a contextual view on problem solving, [3,4,7,8,47]. The central idea is that a problem is determined by constraints that are modified as the problem-solving process proceeds. The constraints can be of three sorts: conditions on the solution, methodological instructions (including heuristic instructions as well as examples of successful problem solutions), and (contextual) certainties (which form the so-called conceptual frame). Some handy sources for the logical aspects: [12,19,20,27] for adaptive logics, [17,30] for the prospective dynamics, and [65,66,67,68] for erotetic logic.

2. Formal Problem-Solving Processes

A fpsp is a chain of stages and a stage (of a fpsp) is a sequence of lines that are written according to a set of instructions: rules with a deontic condition attached to them. The first stage consists of a single line. A subsequent stage is obtained by adding one line to the previous one. The instructions determine which lines may be added at a stage. They obviously depend on the logic or logics of the fpsp.

A fpsp contains problem lines as well as declarative lines. A problem is a non-empty set of questions. In a first approach, one may take the *main* problem to be a single yes–no question $?\{M, \sim M\}$, and all derived problems to be sets of yes–no questions. A and $\sim A$ are the *direct answers* of the question $?\{A, \sim A\}$.

Γ will always denote the premise set. The formula of a declarative line may be a member of Γ or a formula derived from members of Γ. Such formula may have the form

$$[B_1,...,B_n]\,A$$

which indicates that $\Gamma \cup \{B_1,...,B_n\} \vdash A$. The set $\{B_1,...,B_n\}$ will be called the *condition* of A. A formula of the form A will be said to occur unconditionally. However, a formula A is also identified with $[\varnothing]\,A$, whence the condition of A is said to be empty.

Given a problem, the problem solver has to chose a *target* (*from* the problem): a direct answer of a question of the problem. A target is noted on a *target line*; it is a formula the problem solver tries to establish. A target A will be written as

$$[A]\,A$$

which is logically redundant but guides the procedure, as we shall see. If $[A]\,A$ occurs on an unmarked line, A will be said to be an *unmarked target*. The procedure is moreover guided by *marking definitions*. Some marks indicate that a line became useless for solving the main problem; others that some action has to be undertaken to further the solution.

We need some preparatory definitions for the formal machinery. Let $*A$ denote the 'complement' of A, viz. B if A has the form $\sim B$ and $\sim A$ otherwise. Let us distinguish *a*-formulas from *b*-formulas, varying on a theme from [63], and assign to each formula two other formulas according to the following table:

a	a_1	a_2	b	b_1	b_2
$A \wedge B$	A	B	$\sim(A \wedge B)$	$*A$	$*B$
$A \equiv B$	$A \supset B$	$B \supset A$	$\sim(A \equiv B)$	$\sim(A \supset B)$	$\sim(B \supset A)$
$\sim(A \vee B)$	$*A$	$*B$	$A \vee B$	A	B
$\sim(A \supset B)$	A	$*B$	$A \supset B$	$*A$	B
$\sim\sim A$	A	A			

The *positive part* relation is defined recursively by the following clauses:

1. $\mathrm{pp}(A, A)$.
2. $\mathrm{pp}(A, a)$ if $\mathrm{pp}(A, a_1)$ or $\mathrm{pp}(A, a_2)$.
3. $\mathrm{pp}(A, b)$ if $\mathrm{pp}(A, b_1)$ or $\mathrm{pp}(A, b_2)$.
4. If $\mathrm{pp}(A, B)$ and $\mathrm{pp}(B, C)$, then $\mathrm{pp}(A, C)$.

The *instructions* rely on the prospective rules presented in [30]. A fpsp starts with an application of Main, which introduces the main problem. In the instructions, k denotes a suitable line number.

Main Start a fpsp with the line:

$1 \quad \{?\{M, \sim M\}\}$ Main

Target If P is the problem of an unmarked problem line, and A is a direct answer of a member of P, then one may add:

$k \quad [A]A$ Target

Prem If A is an unmarked target, $B \in \Gamma$, and $\mathrm{pp}(A, B)$, then one may add:

$k \quad B$ Prem

The formula analysing rules of **CL** (see [30]) may be summarized as follows:[2]

$$\frac{[\Delta]a}{[\Delta]a_1 \quad [\Delta]a_2} \qquad \frac{[\Delta]b}{[\Delta, {*}b_2]b_1 \quad [\Delta, {*}b_1]b_2}$$

The general form of the rules is $[\Delta]A / [\Delta \cup \Delta']B$. The following instruction handles their application:

FAR If C is an unmarked target, $[\Delta]A$ is the formula of an unmarked line i, $[\Delta]A/[\Delta \cup \Delta']B$ is a formula analysing rule, and $\mathrm{pp}(C, B)$, then one may add:

$k \quad [\Delta \cup \Delta']B \quad i$; R

in which R is the name of the formula analysing rule.

The condition analysing rules of **CL** are summarized by:

$$\frac{[\Delta \cup \{a\}]A}{[\Delta \cup \{a_1, a_2\}]A} \qquad \frac{[\Delta \cup \{b\}]A}{[\Delta \cup \{b_1\}]A \quad [\Delta \cup \{b_2\}]A}$$

2 The left rule states that both $[\Delta]a_1$ and $[\Delta]a_2$ may be derived (separately) from $[\Delta]a$ (notational abuse here and in the text).

They all have the form $[\Delta \cup \{B\}]A$ / $[\Delta \cup \Delta']A$. The following instruction refers to them:

CAR If A is an unmarked target, $[\Delta \cup \{B\}]A$ is the formula of an unmarked line i, and $[\Delta \cup \{B\}]A/[\Delta \cup \Delta']A$ is a condition analysing rule, then one may add:

k $[\Delta \cup \Delta']A$ i; R

in which R is the name of the condition analysing rule.

The instructions EM (excluded middle) and EM0 allow one to eliminate certain problems without answering them.

EM0 If $[\Delta \cup \{*A\}]A$ is the formula of a line i that is neither R-marked nor I-marked, then one may add:

k $[\Delta]A$ i; EM0

EM If A is an unmarked target, $[\Delta \cup \{B\}]A$ and $[\Delta' \cup \{\sim B\}]A$ are the respective formulas of the unmarked or only D-marked lines i and j, and $\Delta \subseteq \Delta'$ or $\Delta' \subseteq \Delta$, then one may add:

k $[\Delta \cup \Delta']A$ i, j; EM

Transitivity is essential for eliminating solved questions as well as for summarizing the remaining problems (and paths) in a problem-solving process.

Trans If A is an unmarked target, and $[\Delta \cup \{B\}]A$ and $[\Delta']B$ are the respective formulas of the at most S-marked[31] lines i and j, then one may add:

k $[\Delta \cup \Delta']A$ i, j; Trans

The last instruction handles derived problems:

3 A line is at most S-marked iff (if and only if) it is not R-marked, not I-marked and not D-marked.

DP If A is an unmarked target from problem line i and $[B_1,...,B_n]A$ is the formula of an unmarked line j, then one may add:

k $\{?\{B_1,\sim B_1\},...,?\{B_n,\sim B_n\}\}$ i, j; DP

In view of the intended applications (deriving predictions, explanations, etc.) the procedural system has no instruction for applying EFQ—the paraconsistent procedural variant logic of **CL** is studied in [22,23].

I now list the *marking definitions*. We shall need several kinds of marks, which have distinct effects on the procedure. Each kind is governed by a definition, which applies to the stages of the fpsp—marks may come and go with each new stage. An R-mark indicates that a line is *redundant*.

Definition 1. *An at most S-marked declarative line* i *that has* $[\Delta]A$ *as its formula is* R-marked *at a stage iff, at that stage,* $[\Theta]A$ *is the formula of a line for some* $\Theta\subset\Delta$.
An unmarked problem line i *is* R-marked *at a stage iff, at that stage, a direct answer* A *of a question of line* i *is the formula of a line.*

Remember that a target is a target *from* a problem line (and from its problem). If A is a target, every line in which $[\Delta]A$ is derived for some $\Delta\neq\{A\}$ will be called a *resolution* line (*for* target A). The line called j in instruction DP will be said to *generate* the problem line introduced by DP. A is *a direct target from* $[\Delta]B$ iff $[\Delta]B$ is the formula of the resolution line that generates problem P, A is a target from P, and $A\in\Delta$—note that some targets from P are not members of Δ.[4] A *target sequence* is a sequence $([\Delta^1]A^1,...,[\Delta^n]A^n)$ in which every A^{i+1} $(1\leq i<n)$ is a direct target from $[\Delta^i]A^i$. A target sequence $([\Delta^1]A^1,...,[\Delta^n]A^n)$ is *grounded* iff A^1 is not a direct target from any unmarked $[\Theta]B$ derived in the fpsp. A set is flatly inconsistent iff it contains A as well as $\sim A$ for some A. I-marked lines are *inoperative*: acting on the line is not useful for solving any extant problem.

4 A resolution line containing $[q]p$ leads to the derived problem $?\{p,\sim p\}$ and $\sim p$ is a target from this.

Definition 2. *An at most S-marked target line that has* $[A]A$ *as its formula is* I-marked *at a stage iff every problem line from which* A *is a target is marked at that stage.*
An at most S-marked resolution line of which $[\Delta^1]A^1$ *is the formula for some* $\Delta^1 \neq \varnothing$ *is* I-marked *at a stage iff, at that stage, for every grounded target sequence* $([\Delta^n]A^n,...,[\Delta^1]A^1)$,

(i) *some target* $[A^i]A^i$ *(*$1 \leq i \leq n$*) is marked, or*

(ii) $\{A^n,..., A^1\} \cap \Delta^1 \neq \varnothing$, *or*

(iii) $\Delta^1 \cup ... \cup \Delta^n \cup \Gamma_s^\circ$ *is flatly inconsistent.*

An unmarked problem line is I-marked *iff no unmarked resolution line generates it.*

A is a *dead end* iff A is (in the present propositional context) a literal and A is not a positive part of a premise. If A is not a literal, then CAR leads from the unmarked $[\Delta \cup \{A\}]B$ to one or more $[\Delta \cup \Delta']B$. The latter is called a CAR-descendant of $[\Delta \cup \{A\}]B$. D-marks indicate that no further action can be taken in view of a line.

Definition 3. *An at most S-marked resolution line with formula* $[\Delta]A$ *is* D-marked *at a stage iff some* $B \in \Delta$ *is a dead end or, at that stage, all CAR-descendants of* $[\Delta]A$ *occur in the fpsp and are D-marked.*
An at most S-marked target line with formula $[A]A$ *is* D-marked *at a stage iff* A *is a dead end or no further action can be taken in view of target* A.[5]

The procedure may be *sped up* (in a way that derives from a straightforward insight in the fpsp) by S-marks. Let Γ_s° be the union of the set of premises and the set of unconditional formulas that occur at stage s of the fpsp.

5 Whether an action that can be taken in view of target A is obvious in view of the instructions. If no further action can be taken in view of target A, all instructions that could be applied in view of that target were taken and resulted in lines that are R-marked, I-marked or D-marked.

Definition 4. *A R-unmarked resolution line in which* $[\Delta^1]A^1$ *is derived is* S-marked *iff*

(i) $\Delta^1 \cap \Gamma_s^\circ \neq \varnothing$, *or*

(ii) *for some target sequence* $([\Delta^n]A^n,...,[\Delta^1]A^1)$, $\{A^n\} \cup \Delta^1$ *is flatly inconsistent whereas* Δ^1 *is not flatly inconsistent, or*

(iii) $\Delta_1 \subset \Delta^n \cup ... \cup \Delta^2$ *for some target sequence* $([\Delta^n]A^n,...,[\Delta^1]A^1)$.

What precedes describes a rather permissive procedure in which the only declarative logic is **CL**. Below I illustrate the procedure by means of an extremely simple example. The prospective dynamics for adaptive logics is hardly more complicated—see [17].

Let the premise set be $\{t \supset u, \sim u \vee r, (r \wedge t) \vee s, \sim s, (q \vee u) \supset (\sim t \vee q)\}$ and let the main problem be $?\{p \vee q, \sim (p \vee q)\}$. The superscripts of the marks name the stage at which the line is (thus) marked—this convention saves rewriting.

1	$\{?\{p \vee q, \sim (p \vee q)\}\}$	Main	
2	$[\sim (p \vee q)] \sim (p \vee q)$	Target	D^3
3	$[\sim p, \sim q)] \sim (p \vee q)$	2; C$\sim \vee$E	D^3

As $\sim p$ is not a positive part of any premise, line 3 is D-marked, and hence so is line 2.

4	$[p \vee q] p \vee q$	Target	
5	$[p] p \vee q$	4; C$\vee$E	D^5
6	$[q] p \vee q$	4; C$\vee$E	
7	$\{?\{q, \sim q\}\}$	4, 6; DP	
8	$[q] q$	Target	
9	$(q \vee u) \supset (\sim t \vee q)$	Prem	
10	$[q \vee u] \sim t \vee q$	9; $\supset$E	
11	$[q] \sim t \vee q$	10; C$\vee$E	
12	$[q, t] q$	11; $\vee$E	I^{12}
13	$[u] \sim t \vee q$	10; C$\vee$E	

14	$[u,t]q$	13; $\vee$E	
15	$\{?\{u,\sim u\},?\{t,\sim t\}\}$	8, 14; DP	
16	$[t]t$	Target	
17	$(r \wedge t) \vee s$	Prem	
18	$[\sim s]r \wedge t$	17; $\vee$E	
19	$[\sim s]t$	18; $\wedge$E	S^{22}
20	$\{?\{s,\sim s\}\}$	16, 19; DP	R^{22}
21	$[\sim s]\sim s$	Target	R^{22}
22	$\sim s$	Prem	

The S-mark of line 19 indicates that Trans can be applied. I repeat part of the fpsp:

14	$[u,t]q$	13; $\vee$E	S^{23}
15	$\{?\{u,\sim u\},?\{t,\sim t\}\}$	8, 14; DP	R^{23}
16	$[t]t$	Target	R^{23}
17	$(r \wedge t) \vee s$	Prem	
18	$[\sim s]r \wedge t$	17; $\vee$E	
19	$[\sim s]t$	18; $\wedge$E	S^{22} R^{23}
20	$\{?\{s,\sim s\}\}$	16, 19; DP	R^{22}
21	$[\sim s]\sim s$	Target	R^{22}
22	$\sim s$	Prem	
23	t	19, 22; Trans	

Again, the S-mark of line 14 indicates how to proceed.

24	$[u]q$	14, 23; Trans	S^{29} R^{30}
25	$\{?\{u,\sim u\}\}$	8, 24; DP	R^{29}
26	$[u]u$	Target	R^{29}
27	$t \supset u$	Prem	
28	$[t]u$	27; $\supset$E	S^{28} R^{29}
29	u	23, 28; Trans	
30	q	24, 29; Trans	

Line 14 is R-marked at stage 24. The S-mark of line 24 triggers Trans, which results in 30. Several marks are added at stage 30, most importantly, line 6 is S-marked, which leads to

31	$p \vee q$	6, 30; Trans

whence the problem is solved (and line 1 is R-marked).

The example does not present the shortest way to derive an answer to $?\{p \vee q, \sim (p \vee q)\}$ from the premise set. Rather, it illustrates most of the features of (these simple) fpsps. Most of the heuristics is pushed into the fpsp and each step of the fpsp is sensible in view of the previous stage. Note that several premises were not introduced and cannot be introduced into the fpsp for this main problem.

3. Adaptive Logics

Here is a loose characterization: a logic is adaptive iff it adapts itself to the specific premises to which it is applied. The *logic* adapts itself to the premises: it depends on the premises which instances of inference rules are correct. And the logic adapts *itself* to the premises: the reasoner does not interfere in this.

The first studied adaptive logics were inconsistency-adaptive—see [5,6,9]. As a growing multiplicity of adaptive logics were described, the need for systematization presented itself. The idea was to find a common formal characterization, which was called *the standard format*. As the basic mechanism behind all adaptive logics is the same, the standard format should do most if not all of the work. It does.

In the sequel, I shall consider only adaptive logics in standard format. I shall try to avoid to complex technicalities, for which I refer to other papers, but rather try to clarify the way in which adaptive logics characterize defeasible reasoning processes.

The standard format is both simple and perspicuous. It was first presented in [12] and was studied more thoroughly in [19]. An adaptive logic **AL** is in standard format if it is characterized by a triple consisting of the following elements:

(i) **LLL**, a lower limit logic,
(ii) Ω, a set of abnormalities that all have the same logical form,
(iii) an adaptive strategy.

Each of these elements has a specific technical function.

The lower limit logic **LLL** determines the part of the adaptive logic that is not subject to adaptation. From a proof theoretic point of view, **LLL** delineates the rules of inference that hold unexceptionally. From a semantic point of view, the adaptive models of Γ are a selection of the **LLL**-models of Γ. It follows that $Cn_{\mathbf{LLL}}(\Gamma) \subseteq Cn_{\mathbf{AL}}(\Gamma)$. In principle, the lower limit logic is a monotonic and compact logic.

Abnormalities are formulas that are presupposed to be false, unless and until proven otherwise. "Abnormality" is a technical term. The abnormalities of corrective adaptive logics are logical falsehoods of **CL**, but the abnormalities of ampliative adaptive logics are not. Thus, in a context in which abductive reasoning is appropriate, it will be supposed that one of the known possible 'causes' of a fact is indeed the cause of this fact. Similarly, in a context in which inductive generalization is appropriate, the presence of instances of $Px \wedge Qx$ and the absence of instances of $Px \wedge \sim Qx$ will be taken a reason to consider $\exists x(Px \supset Qx)$ as true—there are some complications, but these need not concern us here.

Ω comprises all formulas of a certain logical form, which may be restricted. For many inconsistency-adaptive logics, Ω is the set of formulas of the form $\exists(A \wedge \sim A)$, the existential closure of $(A \wedge \sim A)$. For other inconsistency-adaptive logics, the set is restricted, for example, to formulas in which A is a primitive formula—a formula that contains no logical symbols except for identity. Similar restrictions are imposed on many ampliative adaptive logics. See, for example, [26] and [25]. Where introduced, the restriction is justifiable and desirable—if it were not introduced, a flip-flop logic would result.[6] Examples of such inconsistency-adaptive logics are those that have

6 A flip-flip logic is an extremely simple adaptive logic **AL** for which $Cn_{\mathbf{AL}}(\Gamma)$ is identical to $Cn_{\mathbf{LLL}}(\Gamma)$ if Γ requires some abnormalities to be true and is identical to $Cn_{\mathbf{ULL}}(\Gamma)$ if Γ requires no abnormality to be true—see below in the text on **ULL**. Most flip-flop logics have no sensible applications

as their lower limit logic, for example, Schütte's $\mathbf{\Phi}_v$ from [61] (called **CLuNs** in Ghent—see [24]), Priest's **LP** from [58], or Meheus' **AN**$\varnothing$ from [44], and modal adaptive logics that characterize paraconsistent inference relations under a translation, for example [46,49,50]. Incidentally, if the logical form characterizing $\mathbf{\Omega}$ is restricted, every formula of the unrestricted form entails a disjunction of formulas of the restricted form.

To illustrate the variety of sets of abnormalities, I list just a few other logical forms that characterize, possibly with a restriction, the set of abnormalities of some adaptive logics: $\exists A \wedge \exists \sim A$, A, $\Diamond A \wedge \sim A$, and $(A \supset B) \wedge (A \wedge \sim B)$.[7]

That $\mathbf{\Omega}$ is characterized by a logical form is not unimportant. It enables one to consider adaptive logics as formal logics. They are defined by **LLL**, which is a formal logic in the standard sense, and by the supposition that all formulas of a certain logical form are false until and unless proven otherwise.[8]

Extending the lower limit logic with the requirement that no abnormality is logically possible, results in the *upper limit logic* **ULL**. Syntactically, **ULL** is obtained by extending **LLL** with an axiom stating that members of $\mathbf{\Omega}$ entail triviality. Semantically, the upper limit logic is characterized by the lower limit models that verify no abnormality. **ULL** requires premise sets to be normal, and 'explodes' abnormal premise sets (assigns the trivial consequence set to them).

So **ULL** is the logic that is appropriate for premise sets that do not require any abnormalities to hold true. For corrective adaptive logics have a lower limit logic that is weaker than **CL**; most of them have **CL** as their upper limit logic. One normally would apply **CL**, but if the premises do not allow this (because their **CL**-consequence set is trivial), the adaptive logic

7 The abnormalities are the formulas that have the (possibly restricted) form and are **LLL**-contingent.

8 This is one of the central differences with the so-called formula-preferential systems from [2,37], which refer to an arbitrary set of formulas of which as many members as possible are considered as true. See also [27] for the relation between adaptive logics and formula-preferential systems.

interprets the premise set as much as possible according to **CL**. Ampliative adaptive logics have an upper limit logic that has often no sensible application contexts itself. This does not make the adaptive logic meaningless. For example, we all know that the world we live in is not completely uniform: not all objects have the same properties. Nevertheless, it makes sense to interpret the world 'as uniform as possible'. The logic of inductive generalization does precisely this: it takes the world to be as uniform as the empirical data permit.

Although corrective adaptive logics have a lower limit logic that is weaker than **CL**, it is possible to add new logical symbols that have exactly the same meanings as the original symbols of **CL**. These new connectives are extremely helpful from a technical point of view, but obviously do not occur in the premise set. They are means to handle the premises, to interpret them as normally as possible, even if the premises require abnormalities to hold true.

As was suggested before, if the premise set does not require any abnormality to obtain, then the adaptive logic delivers the same consequences as the upper limit logic. If the premise set requires some abnormalities to obtain, the adaptive logic will still deliver more consequences than the lower limit logic, viz. all upper limit consequences that are not 'blocked' by those abnormalities.[9] In sum, the adaptive logic interprets the set of premises 'as normally as possible'; it takes abnormalities to be false 'in as far as' the premises permit.

An *adaptive strategy* is required because many premise sets **LLL**-entail a disjunction of abnormalities (members of Ω) without entailing any of its disjuncts. Disjunctions of abnormalities will be called *Dab-formulas*. In the sequel, any expression of the form $Dab(\Delta)$ will refer to the (classical) disjunction of the members of a finite $\Delta \subseteq \Omega$. *Dab*-formulas that are derivable by the lower limit logic from the premise set Γ will be called *Dab-consequences* of Γ. If $Dab(\Delta)$ is a *Dab*-consequence of Γ, then so is $Dab(\Delta \cup \Theta)$ for any finite $\Theta \subset \Omega$. For this reason, only the *minimal Dab*

9 Flip-flop logics are an exception—see footnote 6.

-consequences of the premise set are relevant: $Dab(\Delta)$ is a *minimal Dab*-consequence of Γ iff $\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta)$ and there is no $\Theta \subset \Delta$ such that $\Gamma \vdash_{\mathbf{LLL}} Dab(\Theta)$. If $Dab(\Delta)$ is a minimal *Dab*-consequence of Γ, then Γ determines that some member of Δ behaves abnormally, but fails to determine which member of Δ behaves abnormally. We have seen that adaptive logics interpret a premise set 'as normally as possible'. As some minimal *Dab*-consequences of Γ may contain more than one disjunct, this phrase is ambiguous. It is disambiguated by choosing a specific adaptive strategy.

Reliability from [6] is the oldest known strategy. The minimal abnormality strategy, first presented in [5], delivers at least the same consequences as the Reliability strategy. For some lower limit logics and sets of abnormalities Δ is a singleton whenever $Dab(\Delta)$ is a minimal Dab-consequence of a premise set. If this is the case, the Reliability strategy and the Minimal Abnormality strategy lead to the same result and coincide with what is called the *Simple* strategy—see [26,44,45] for examples. Most other strategies were needed to characterize an existing consequence relation by an adaptive logic—see [11,14,28,34,64] for examples. In this paper, I only mention the marking definition of the Reliability strategy, referring to [19] for the marking definitions of the Minimal Abnormality strategy and of the Simple strategy (and to the specific papers for the making definitions of the other strategies).

Adaptive logics handle defeasible inferences. Some formulas derived at a point have to be considered as OUT at a later point, and possibly as IN at a still later point. This dynamics should be controlled. The means to do so in the dynamic proofs of adaptive logics are the marking definitions. They are phrased in terms of the conditions that are attached to lines of annotated proofs.[10]

A line of an annotated dynamic proof consists of a line number, a formula, a justification, and a (possibly empty) condition. The proofs are governed by three (generic) rules and a marking definition. The rules, which depend on $\mathbf{LLL}$ and Ω only, determine which lines may be added to the

10 The definitions may be adjusted for non-annotated proofs, which are just lists of formulas as usual.

proof; the marking definitions, which depend on Ω and the strategy, determine which formulas are IN or OUT at some stage of the proof. Let

$A \quad \Delta$

abbreviate that A occurs in the proof on the condition Δ, the rules may then be phrased as follows:

PREM If $A \in \Gamma$:

$$\frac{\begin{array}{cc} \dots & \dots \end{array}}{\begin{array}{cc} A & \varnothing \end{array}}$$

RU If $A_1, \dots, A_n \vdash_{\mathbf{LLL}} B$:

$$\frac{\begin{array}{cc} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \end{array}}{\begin{array}{cc} B & \Delta_1 \cup \dots \cup \Delta_n \end{array}}$$

RC If $A_1, \dots, A_n \vdash_{\mathbf{LLL}} B \vee Dab(\Theta)$

$$\frac{\begin{array}{cc} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \end{array}}{\begin{array}{cc} B & \Delta_1 \cup \dots \cup \Delta_n \cup \Theta \end{array}}$$

Only the conditional rule, RC, introduces non-empty conditions. The unconditional rule, RU, simply carries the conditions over from given formulas to formulas that are derived from them.

The marking definitions require some preparation. $Dab(\Delta)$ is a *minimal* Dab*-formula* at stage s of the proof if, at stage s, $Dab(\Delta)$ occurs in the proof on the empty condition and, for any $\Delta' \subset \Delta$, $Dab(\Delta')$ does not occur in the proof on the empty condition. Where $Dab(\Delta_1)$, …, $Dab(\Delta_n)$ are the minimal Dab-formulas at stage s of the proof, $U_s(\Gamma) = \Delta_1 \cup \dots \cup \Delta_n$ is the set of unreliable formulas at stage s. So, whenever $Dab(\Delta)$ is a *minimal* Dab-formula at stage s, all members of Δ are considered as unreliable at that stage (at least one of them is true, but it is not determined which one). Further Dab-formulas may be derived at a later stage s'. So $U_{s'}(\Gamma)$ may

comprise abnormalities that are not a member of $U_s(\Gamma)$. The opposite is also possiblen viz. if $Dab(\Delta)$ is derived at stage s and, for some $\Theta \subset \Delta$, $Dab(\Theta)$ is derived at stage s'. A line is marked (for Reliability) at stage s iff a member of its condition is unreliable at that stage:

Definition 5. Marking for Reliability: *Line i is marked at stage s iff, where Θ is the condition of line i, $\Theta \cap U_s(\Gamma) \neq \emptyset$.*

A formula is *derived* from Γ *at a stage* of the proof iff it is the formula of a line that is unmarked at that stage. As the proof proceeds, unmarked lines may be marked and vice versa. So, it is important that one defines a different, stable, kind of derivability:

Definition 6. *A is* finally derived *from Γ on line i of a proof at stage s iff (i) A is the formula of line i, (ii) line i is not marked at stage s, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.*

A different definition goes as follows: A is *finally derived* on line i of a (possibly infinite) proof from Γ iff i is unmarked and the proof is *stable* with respect to line i, viz. line i is unmarked in all extensions of the proof. The previous definition is more appealing. The only way to establish the existence of a proof in which A is finally derived is by a metalinguistic reasoning anyway. Moreover, the definition has a nice game-theoretic interpretation: if an opponent is able to extend the proof in such a way that line i is marked, then the proponent is able to extend it further in such a way that line i is unmarked.

Definition 7. $\Gamma \vdash_{\mathbf{AL}} A$ (*A is* finally **AL**-derivable from Γ*) iff A is finally derived on a line of a proof from Γ.*

To describe the *semantics*, let $M \models A$ denote that M assigns a designated value to A, in other words that M verifies A, and let the abnormal part of a **LLL**-model M be defined as follows:

Definition 8. $Ab(M) = \{A \in \Omega \mid M \models A\}$

Where $Dab(\Delta_1), Dab(\Delta_2), \ldots$ are the minimal minimal Dab-formulas that are verified by all **LLL** models of Γ, $U(\Gamma) = \Delta_1 \cup \Delta_2 \cup \ldots$ is the set of formulas that are unreliable with respect to Γ.[11] Let $\mathbf{AL}^r$ and $\mathbf{AL}^m$ be the adaptive logics defined by a **LLL**, an Ω, and the Reliability strategy, respectively the Minimal Abnormality strategy.

Definition 9. *A* **LLL***-model* M *of* Γ *is* reliable *iff* $Ab(M) \subseteq U(\Gamma)$.

Definition 10. $\Gamma \models_{\mathbf{AL}^r} A$ *iff* A *is verified by all reliable models of* Γ.

Definition 11. *A* **LLL***-model* M *of* Γ *is* minimally abnormal *iff there is no* **LLL***-model* M' *of* Γ *such that* $Ab(M') \subset Ab(M)$.

Definition 12. $\Gamma \models_{\mathbf{AL}^m} A$ *iff* A *is verified by all minimally abnormal models of* Γ.

Note that the proof theory as well as the semantics of an adaptive logic are fixed by the standard format. There is more. Many metatheoretic properties of adaptive logics can be proved from the format itself, rather than from the specific properties of the logic. Among such properties are: the *Soundness* and *Completeness* proof (given that the lower limit logic is sound and complete with respect to its semantics), the *Derivability Adjustment* Theorem ($\Gamma \vdash_{\mathbf{ULL}} A$ iff there is a finite $\Delta \subset \Omega$ such that $\Gamma \vdash_{\mathbf{LLL}} A \vee Dab(\Delta)$), *Proof Invariance* (a formula finally derived in some proof from Γ can be finally derived in every proof from Γ), *Strong Reassurance* (if M is a **LLL**-model of Γ but not an **AL**-model of Γ, then there is an **AL**-model M' such that $Ab(M') \subset Ab(M)$), Cautious Cut, Cautious Monotonicity, and many further properties—see especially [19].

Nearly all known inconsistency-adaptive logics have a characterization in standard format.[12] In some cases, forging a consequence relation into standard format may require a translation, for example to a modal language. Even where it is useful to provide 'direct dynamic proofs' in untranslated terms (see [32] or [48] for examples) the formulation in standard format

11 So $U(\Gamma)$ is like $U_s(\Gamma)$, except that it refers to final derivability (or rather to its semantic counterpart) and not to a stage of a proof.

12 An exception is Priest's $\mathbf{LP}^m$ from [59] and emended in [60]. This adaptive logic proceeds in terms of properties of the model, rather than in terms of the formulas verified by the model.

has the advantage to provide the proof theory, semantics and metatheoretic properties, and to warrant (by an easy demonstration) that the direct proof theory is correct.

In many situations one needs to combine adaptive logics. This too may be handled in a generic way. Two general stratagems for combining adaptive logics were designed. Apparently they serve all needs. Again, the combination stratagems, rather than specific properties of the combined logics, warrant that the combination does the desired job. The matter is studied, for example, in [12].

4. Some Extra-Logical Extensions

Some questions can be directly answered by standard means, for example questions that are answered (obviously at a cost) by observational or experimental means. Building this into fpsps is a bit more tricky than is usually thought (and than I supposed before). Hintikka, for example in [35], tries to handle this in terms of a set of formulas. If a direct answer to a question is a member of this set, and the question is posed, then the answer is provided by the oracle. This will not do because some questions are only answered in specific situations and this cannot be expressed by means of **CL**-formulas. Consider an example.

Let p be "the shutter of the window is open" and let q be "the sun shines". Suppose that I know that the position of the sun is such that, if the shutter is open and the sun shines, then I see a spot of sunlight on the floor, but that I do not know (and cannot observe in the present situation) whether the shutter is open and that I do not know (and cannot observe in the present situation) whether the sun is shining. So, in the present situation, the oracle informs me that $\sim p \vee \sim q$ (either the shutter is closed or the sun does *not* shine). Suppose, moreover, that the sun is actually shining, p, and that the shutter of the window is closed. So, if the shutter were open, I would see that the sun is shining. The trouble is that this cannot be expressed by any statement in the oracle set. Indeed, the only candidate seems to be $\sim p \vee q$. But if that were in the oracle set, I could force the oracle in the present situation to answer both $\sim p \vee \sim q$ and $\sim p \vee q$, from which $\sim p$

follows. However, the oracle cannot provide me with this information in the present situation (we supposed that I cannot observe in the present situation whether the shutter is open). So if plain formulas go into the oracle set, its contents has to vary with changes in the situation, contrary to Hintikka's implicit setup. The proposal I made in [13], viz. that the set contains answerable questions, is affected by the same problem.[13]

How should the oracle be handled? The researcher is convinced that a certain question Q is answered if certain circumstances Δ obtain. Let this belief be expressed by $(\Delta : Q)$. In the example from the previous paragraph, the formula is $(\{p\} : ?\{q, \sim q\})$. Such a formula, respectively statement, may be handled as a premise.

Obviously, the statement $(\Delta : Q)$ may be false. But if the researcher is convinced of its truth (which may be seen from its occurrence in the fpsps) and if he or she is convinced that all members of Δ hold true (because they occur in the fpsp), then he or she will *take* his or her observation as an answer to the question. Consider again the example from two paragraphs ago. Suppose I press a button which I believe to open the shutter and that I do not see a spot of sunlight thereafter. I will then conclude that the sun is not shining. If the button did not function and the shutter is still closed, I shall still believe that the sun is not shining, because I believe the shutter to be open.

So the oracle is an epistemic device, not an ontological one. Formally, it may be represented by a function that maps couples composed of a question and a set of statements to a direct answer of the question, for example $f(Q, \Delta) = A$, in which A is a direct answer to Q. If $(\Delta : Q)$ and all members of Δ occurs in the fpsp, the researcher obtains one of the direct answers of Q. So the connected instruction is as follows:

13 That $\sim p \vee q$ cannot be put into the set is a result of the fact that it does not express the intended counterfactual or conditional.

New If A is an unmarked target, $\mathrm{pp}(A, B)$ for some direct answer B of Q, $(\Delta : Q)$ and all members of Δ occur in the fpsp, then one should add, for some direct answer C of Q:

k C i; New

Of course, the researcher does not know beforehand which new premise will be introduced if $(\Delta : Q)$ and the members of Δ occur in the fpsp. So the heuristics cannot proceed in terms of the answers that would be provided if the conditions for New were realized, but only in terms of the statements of the form $(\Delta : Q)$ that are held true. Precisely such statements are essential for setting up experiments and for interpreting their outcome. It is instructive to reconstruct for example Galilei's inclined plane experiment in terms of such knowledge. Planning an experiment may be seen as itself resulting from a fpsp.

I have supposed that a new premise is provided as soon as $(\Delta : Q)$ and all members of Δ occur in the fpsp. One might object that posing the question to the oracle, viz. actually making the relevant observation, may come at some expense, and hence that it should be left to the researcher's decision whether the question Q is actually introduced. To modify the formal machinery accordingly is clearly not difficult, but I take it to be superfluous. Indeed, the expense might be associated to the members of the condition Δ in $(\Delta : Q)$. If the shutter is open and if I turn my head in the right direction and if I pay the required attention to the presence of a spot of sunlight on the floor, then I have the answer to the question whether the sun is shining.

The expense for realizing certain statements to hold is obviously an important factor. If the expense for answering the main problem by direct means is near to nil, to do so would be preferred over engaging in logical deduction. If the expense is extremely high, one may renounce from seeking a direct answer, even where this entails that the main problem is left unsolved. However, such considerations will be neglected in the present context.

The presence of $(\Delta : Q)$ and of the members of Δ in the fpsp triggers an application of New as soon as some unmarked target A is a positive part of a direct answer of Q. If there is no such target A, the answer to Q is,

for all one knows, irrelevant to the extant problems and hence should not be introduced. However, this should be refined in future work. Suppose that Q is $?\{B, \sim B\}$, that B is present in the fpsp, but that we observe $\sim B$. This reveals that there is a problem with the fpsps, and hence that it has to be restructured in order to solve the problem (possibly by introducing an inconsistency-adaptive logic).

The present extension requires that one redefines: A is a *dead end* iff A is (in the present propositional context) a literal and A is not a positive part of a premise or of a direct answer to a question Q which is such that $(\Delta : Q)$ is among the premises for some Δ. It also requires an instruction for introducing premises of the form $(\Delta : Q)$. This instruction should obviously require that $(\Delta : Q)$ is only introduced if some unmarked target is a positive part of a direct answer to Q. Finally, handling the matter adequately requires that fpsps are extended with a feature indicating that the researcher performs a certain action in an attempt to realize that the members of Δ are true, and that he or she reasons about such actions, taking their cost into account. Nothing of this involves serious difficulties. Once it is built into the fpsps, the latter will be able to guide action and the ensuing observations.

A second extralogical extension was announced, but this is easier to handle. It would be very unrealistic to suppose that a problem-solving process proceeds from all available knowledge. It actually proceeds from a set of knowledge that is judged relevant by the researcher and that he or she is sufficiently acquainted with. Apart from the knowledge that forms the premise set, lots of knowledge is available within the scientific community. Some of it may turn out relevant as the problem-solving process proceeds, for example because the problem may be suspected not to be solvable from the present premise set and because some targets are positive parts of available knowledge.

If a body of available knowledge is added to the premise set, a researcher may need to study this knowledge extensively or may invoke the help of a specialist in the domain. The reasons for this are double. On the one hand, a specialist will know many consequences of the body of knowl-

edge which are directly relevant for the extant problem, whereas it may take the researcher lots of (logical) work to see what follows and what of this is relevant. Moreover, the specialist will be able to determine which changes should be made to a fpsp in order to accommodate the new body of knowledge. The changes may pertain to different elements of the fpsp: the logics, the procedure, the heuristic instructions, and the questions that are answerable by standard means.

5. On Heuristics

In [35] and elsewhere, Hintikka draws a sharp distinction between the rules of the logic and the heuristics by which a desired proof is obtained—I avoid Hintikka's "strategy" because of the technical sense of this term in Section 3. Hintikka's favoured comparison is with games of chess: the rules determine the possible moves whereas the heuristics is directed towards applying the rules in such a way that the game is won. According to this view, the forms of reasoning that are explicated by adaptive logics are not a matter of logic at all, but a matter of heuristics and hence should be explicated in heuristic terms.

The least one can say is that the comparison does not add much bite to the distinction. For one thing, the normal outcome of a game of chess is a checkmate king, whereas a proof exhibits a reasoning that leads from (a subset of) the premises to the conclusion. So, in the case of proofs, unlike in the case of chess, both the heuristics and its result (even according to the standard view) are of the same kind, viz. concern reasoning. That the heuristics concerns reasoning clarifies at once why it is possible, for example in terms of the prospective dynamics, to push part of the heuristics into the proof itself. In other words, the bookkeeping of part of the heuristic reasoning, which is usually done outside of the proof, can very well be done within the proof.

Prospective proofs still require a heuristics. However, as a result of their proof format, the delineation between rules and heuristics lies at a different point than with more usual proofs. The matter is even modified more drastically if prospective proofs are phrased in terms of a procedure—remember

that this is composed by instructions: rules with permissions and obligations attached to them. The more restrictive the deontic qualifications of the instructions, the less room is left for heuristic considerations. Note that the extreme case, viz. a deterministic procedure, is by no means a universal ideal. The room that should be reserved for the heuristics depends on the specific application.

The difference of opinion with Hintikka has a principled side and a pragmatic one. From a pragmatic point of view, I see it as an advantage of my approach that part of the usual heuristics is fixed by the chosen adaptive logics, the prospective dynamics, and the (more or less restrictive) procedure by which the latter is governed. These three elements provide means to articulate those parts of the heuristics that are incorporated into them. Once articulated, they can be evaluated, discussed, and replaced where necessary. In other words, it is a pragmatic advantage of my view that the claim that "it is a matter of heuristics" can be replaced by definite and precise statements about an articulated framework.[14]

However, there was also a principled side. We have seen above that the delineation between rules and heuristics is not a fixed one. It is instructive to see why this is so. Consider again Hintikka's view. In view of a given purpose, a specific heuristics is applied in the framework of $\mathbf{CL}$. For every premise set Γ (possibly combined with a problem, say to find an explanation for Pb), there will be a set $f(\Gamma)$ of formulas that can be obtained by the heuristics. So the heuristics can be seen as a function that maps every set Γ to a set $f(\Gamma)$. So it is a logic, a mapping $\wp(W) \mapsto \wp(W)$ (in which W is the set of closed formulas and $\wp(W)$ its power set). Obviously, this logic may have a number of unusual properties; it may for example be non-monotonic, or even non-reflexive. But is is a logic in the above sense. So it makes sense to study its proof theory and to use it as an explication for the connected form of reasoning, and it makes sense to study its semantics and the properties of its models. One may still try to argue that the resulting logic is not the

14 Needless to say, there still is a need to make more precise claims about the heuristics that remains once the adaptive logics, their prospective dynamics and the governing procedure have been fixed.

standard of deduction. I am not sure that one will succeed in establishing that some other logic is the standard of deduction, but I grant that the attempt to do so may be sensible. Yet, even if the logic is not the standard of deduction and some other logic is, it still is a logic and this was the point I had to make.

6. On Content Guidance

A first, general, and hence non-specific form of content guidance lies with the language of a scientific discipline and with the connected conceptual system. Both are the result of the development that led to the present state of the discipline. They incorporate a set of presuppositions about the domain. They determine the way in which problems are phrased and tackled.

A clear and specific form of content-guidance resides in the use of premises of the form $(\Delta : Q)$, which was discussed in Section 4. There we have seen that the oracle is actually an epistemic device, referring to current knowledge. Similarly, the fact that the premise set may be extended with available knowledge that was judged irrelevant before, constitutes a clear form of content-guidance.

A point that pertains to all adaptive logics is that these logics validate certain *applications* of rules that transcend their lower limit logic. It depends on the content of the premise set, as determined by the lower limit logic, whether certain consequences are drawn from the premises. This important point deserves to be spelled out clearly. It is not that the consequence set is a function of the premise set—a feature that is shared by nearly all logics. The point is that the logics adapt *themselves* to the premises: whether a conclusion follows from some premises does not merely depend on properties of the logic, but also on the other premises. So this form of content guidance is built in into adaptive logics.

There is a multiplicity of adaptive logics for every purpose. Consider handling inconsistency as an example. Many paraconsistent logics can be taken as the lower limit logic and combined with a suitable set of abnormalities and a strategy. Moreover, some forms of handling inconsistency are

characterized by an adaptive logic under a translation. A good example form the Rescher–Manor mechanisms. They are characterized by an adaptive logic in standard format, but under a translation—see for example [11]. They also have a so-called direct proof theory, as was shown in [32], but this is not connected to an adaptive logic in standard format. The situation is similar for the so-called signed systems for paraconsistent reasoning from [33], as was shown in [28]. The variety is even larger than that. An inconsistency-adaptive logic need not be the most suitable instrument to handle a theory that turned out to be inconsistent. In some cases, it is more fitting to apply a logic that is adaptive with respect to properties of a different logical symbol. This point was already made in [10] and a case study is forthcoming in [31].

Given the multiplicity of adaptive logics for nearly every purpose, the choice has to be justified. The justification can only be obtained by logically analysing the concrete case one is confronting—see again [31] for an example. In other words, the choice of the adaptive logics that should be applied in a fpsp is typically content dependent.

The selection of an adaptive logic is typically a *local* matter. This means that the result of a fpsp will lead to certain changes in the setup of the fpsp. Consider the case where one tries to obtain an explanation for a certain fact. If one hopes that a theory is able to provide the explanation, the suitable way to handle the problem is by a fpsp in which a logic is abduction is set to work. But suppose that, at some point, it becomes dubious whether the theory will provide any explanation. One will then first try to obtain a new generalization that may provide the explanation. So one has to extend the adaptive logics of the fpsp by a logic of inductive generalization. Another example of a local choice occurs when the theory from which one tries to obtain the explanation turns out inconsistent. This insight, possibly gained by the fpsp itself, may require that **CL** is replaced by an inconsistency-adaptive logic as the underlying logic of the theory—see [18] for an analysis of the effect of such a move. In still other cases an invoked theory may prove to be contradicted by available data, and hence need to be introduced as plausible only—there is a multiplicity of adaptive logics handling plausible

premises (see for example [25]). So, again, the choice has to be justified in a contextual way.

The situation is similar for the choice of an erotetic logic: one has to justify the specific choice in terms of the insights provided by the problem-solving process. The situation is also similar for the choice of the specific procedure that governs the prospective dynamics.

A further form of content guidance relates to invoked background theories. In fpsps, background theories can be introduced in a defeasible way (as plausible only) and (as was said before) there is a multiplicity of available means for handling plausible premises. So background theories are taken serious but nevertheless can be rejected or revised where necessary.

It is well-known that the results of scientific discovery and creativity are not fully determined by the situation of the discipline. Even if scientific discovery is seen as a form a problem solving, there should be left ample room for the world-view, personal constraints, …, of individual researchers and research groups—see for example [47]. While it may be difficult to formalize such elements, or even to make them explicit, it is easy enough to take their consequences into account in fpsps. The central feature is that solutions may be arrived at by eliminating certain disjuncts from disjunctions of abnormalities. How this is done is discussed at length in [21] (in the context of inconsistency-adaptive logics) and in [??0] (in the context of inductive generalization), and the mechanism is easily generalized to all adaptive logics. A derived (non-singleton) disjunction of abnormalities leads to the question which of its disjuncts is abnormal. This question can typically be answered, fully or partially, but always in a defeasible way, by analysing the situation in the discipline, but also by world-view considerations, by personal constraints, or even by blind guesses. So fpsps provide a framework that locate the problems, suggest ways out (by clarifying which defeasible statements may provide a solution), and logically guide the handling of such statements. The origin of the statements is obviously extra-logical. This is a form of content-guidance. However, but the logic guides the problem-solving process.

Finally, I come to the heuristics that remains after all other elements of the fpsp have been fixed. There are logical as well as non-logical aspects to

this heuristics. The logical ones are mainly related to the road along which a statement that is derivable by the available means is actually derived. This too is largely a matter of learning. Suppose that one wants to derive theorems from the Peano Axioms. Whatever the target at a specific point in the fpsp, it will always be a positive part of an instance of the axiom of mathematical induction. One will soon find out, however, that the introduction of such instances will only occasionally lead to a successful search path. This insight may lead to the conclusion that one first tries to operate on the other axioms, and only introduces an instance of the induction axiom when that fails. This is just an example out of many. Every logician that has made a set of axiomatic proofs has experienced that each new axiomatic system, even for the same logic, requires another heuristics.

The non-logical aspects are especially related to the kind of move that is invoked at a specific point. Given a target or set of targets, there often is the choice between attempting a theoretical derivation, making an observation, and (devising and) performing an experiment. Which is the right move will have to be determined on the basis of one's insight in the situation and on the analogy between the situation and analogical cases that were successfully solved in the past.

7. Concluding Comments

Let me begin with some warnings. The first is that, in the end, all knowledge is defeasible. It does not follow that all declarative premises of a fpsp should be introduced in a conditional way. This, indeed, would only make sense in a futile attempt to incorporate into the approach all possible outcomes of scientific reasoning—see the comments on the Vienna Circle in the first paragraph. So it cannot and should not be avoided that fpsps are seen as themselves subject to revision. As we have seen in Section 6, the insights provided by a fpsp may lead to changes to the elements of the fpsp. Such changes are justified in view of insights in the situation and in view of insights from the history of science—"what we have learnt about how to learn".

The second warning concerns abduction. There are many ampliative reasoning forms. To reduce all of them to abduction would come to a serious overestimation of the latter and to a simplistic view on scientific discovery and creativity. A further warning concerning abduction itself is appropriate. It is fairly clear what it means that an explanation for a fact is abduced in the presence of a theory—the most recent book on the topic, [1], handles precisely that. A very different matter is an abduction in the absence of a theory. I can only confess that I have read nothing on this that was sufficiently clear to me. Some authors give the impression that such abductions rely on implicit beliefs, which may be connected to action habits rather than to statements. This seems acceptable enough. Only, as the origins of such implicit beliefs are clearly located in our observing and handling of everyday objects, it seems highly unlikely that 'abductions' relying on implicit beliefs could be responsible for creative science or for novel insights.

In this paper, I avoided taking a stand in such philosophical debates as the one opposing realism to positivism. The reason for this is that such debates may be relevant from a general epistemological point of view, but their impact on methodological decisions is often overestimated, at least from a contemporary point of view. Moreover, in as far as they do play a role in problem solving, they may be handled in the same way as world-views or personal constraints. Even if they might occasionally shape a fpsp, the formal framework of fpsps is sufficiently malleable (see Section 6) not to exclude the actual impact of such philosophical stands.

The last warning is that the fpsps approach outlined in the previous sections needs (obviously) further elaboration. Nothing was said, for example, on model-based reasoning. Also, while fpsps demonstrably result in certain conceptual changes, the problem of conceptual change requires a much deeper study. So the approach is far from complete. It was presented here because it is promising and hence deserves further study.

The most important comment is that seeing science as a content-guided problem-solving activity does not exclude a formal approach. Quite to the contrary. Precisely in order to be able to express and understand the content-guidance, one needs a maximally formal approach. A fpsp can be seen

as a frame with many open slots. The slots have to be filled with (declarative as well as erotetic) logics, prospective dynamics, a procedure, and a set of heuristic instructions. The formal framework will not deliver any results unless the slots are filled and to fill them requires a justification. The more structure there is in the entities that may be plugged into the slot, the more it will be possible to make the content-guidance explicit and to justify it. As I see it, this is the great advantage of the fpsps approach over Hintikka's—see Section 5. Rather than keeping the heuristics free from the intrusion of rules and instructions, one has to maximize the intrusion. One has to make an effort to turn all possible aspects of the heuristics into alternative sets of transparent formal entities. Only in doing so, one will be able to make sense of the content-guided aspects of problem solving.

Acknowledgements

Research for this paper was supported by subventions from the Fund for Scientific Research – Flanders, from Ghent University, and from the Royal Flemish Academy of Belgium for Science and the Arts.

References

[1] Atocha Aliseda. *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer, Dordrecht, 2006.

[2] Arnon Avron and Iddo Lev. A formula-preferential base for paraconsistent and plausible non-monotonic reasoning. In *Proceedings of the Workshop on Inconsistency in Data and Knowledge (KRR-4) Int. Joint Conf. on AI (Ijcai 2001)*, pages 60–70, 2001.

[3] Diderik Batens. Incommensurability is not a threat to the rationality of science or to the anti-dogmatic tradition. *Philosophica*, 32:117–132, 1984.

[4] Diderik Batens. Meaning, acceptance, and dialectics. In Joseph C. Pitt, editor, *Change and Progress in Modern Science*, pages 333–360. Reidel, Dordrecht, 1985.

[5] Diderik Batens. Dialectical dynamics within formal logics. *Logique et Analyse*, 114:161–173, 1986.

[6] Diderik Batens. Dynamic dialectical logics. In Graham Priest, Richard Routley, and Jean Norman, editors, *Paraconsistent Logic. Essays on the Inconsistent*, pages 187–217. Philosophia Verlag, München, 1989.

[7] Diderik Batens. Do we need a hierarchical model of science? In John Earman, editor, *Inference, Explanation, and Other Frustrations. Essays in the Philosophy of Science*, pages 199–215. University of California Press, 1992.

[8] Diderik Batens. *Menselijke kennis. Pleidooi voor een bruikbare rationaliteit*. Garant, Antwerpen/Apeldoorn, 1992. 2: 2004.

[9] Diderik Batens. Inconsistency-adaptive logics. In Ewa Orlowska, editor, *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*, pages 445–472. Physica Verlag (Springer), Heidelberg, New York, 1999.

[10] Diderik Batens. A survey of inconsistency-adaptive logics. In Batens et al. [29], pages 49–73.

[11] Diderik Batens. Towards the unification of inconsistency handling mechanisms. *Logic and Logical Philosophy*, 8:5–31, 2000. Appeared 2002.

[12] Diderik Batens. A general characterization of adaptive logics. *Logique et Analyse*, 173–175:45–68, 2001. Appeared 2003.

[13] Diderik Batens. A formal approach to problem solving. In Claudio Delrieux and Javier Legris, editors, *Computer Modeling of Scientific Reasoning*, pages 15–26. Universidad Nacional del Sur, Bahia Blanca, Argentina, 2003.

[14] Diderik Batens. A strengthening of the Rescher–Manor consequence relations. *Logique et Analyse*, 183–184:289–313, 2003. Appeared 2005.

[15] Diderik Batens. The basic inductive schema, inductive truisms, and the research-guiding capacities of the logic of inductive generalization. *Logique et Analyse*, 185–188:53–84, 2004. Appeared 2005.

[16] Diderik Batens. The need for adaptive logics in epistemology. In Dov Gabbay, S. Rahman, J. Symons, and J. P. Van Bendegem, editors, *Logic, Epistemology and the Unity of Science*, pages 459–485. Kluwer Academic Publishers, Dordrecht, 2004.

[17] Diderik Batens. A procedural criterion for final derivability in inconsistency-adaptive logics. *Journal of Applied Logic*, 3:221–250, 2005.

[18] Diderik Batens. The theory of the process of explanation generalized to include the inconsistent case. *Synthese*, 143:63–88, 2005.

[19] Diderik Batens. A universal logic approach to adaptive logics. In Jean-Yves Béziau, editor, *Directions in Universal Logic*. Birkhauser, 2006. In print.

[20] Diderik Batens. *Adaptive Logics and Dynamic Proofs. A Study in the Dynamics of Reasoning*. 200x. Forthcoming.

[21] Diderik Batens. Narrowing down suspicion in inconsistent premise sets. In Jacek Malinowski and Andrzej Pietruszczak, editors, *Essays in Logic and Ontology: dedicated to Jerzy Perzanowski*, Poznań Studies in the Philosophy of the Sciences and the Humanities. Rodopi, Amsterdam, in print.

[22] Diderik Batens. A paraconsistent proof procedure based on classical logic. See http://www.cle.unicamp.br/wcp3/ for an abstract.

[23] Diderik Batens. It might have been Classical Logic. To appear.

[24] Diderik Batens and Kristof De Clercq. A rich paraconsistent extension of full positive logic. *Logique et Analyse*, 185–188:227–257, 2004. Appeared 2005.

[25] Diderik Batens and Lieven Haesaert. On classical adaptive logics of induction. *Logique et Analyse*, 173–175:255–290, 2001. Appeared 2003.

[26] Diderik Batens and Joke Meheus. The adaptive logic of compatibility. *Studia Logica*, 66:327–348, 2000.

[27] Diderik Batens and Joke Meheus. Recent results by the inconsistency-adaptive labourers. In Jean-Yves Béziau and Walter A. Carnielli, editors, *Paraconsistent Logic with no Frontiers*, Studies in Logic and Practical Reasoning. North-Holland/Elsevier, 2006. In print.

[28] Diderik Batens, Joke Meheus, and Dagmar Provijn. An adaptive characterization of signed systems for paraconsistent reasoning. To appear.

[29] Diderik Batens, Chris Mortensen, Graham Priest, and Jean Paul Van Bendegem, editors. *Frontiers of Paraconsistent Logic*. Research Studies Press, Baldock, UK, 2000.

[30] Diderik Batens and Dagmar Provijn. Pushing the search paths in the proofs. A study in proof heuristics. *Logique et Analyse*, 173–175:113–134, 2001. Appeared 2003.

[31] Diderik Batens and Peter Verdée. Content-guidance in the choice of a logic. Forthcoming.

[32] Diderik Batens and Timothy Vermeir. Direct dynamic proofs for the Rescher–Manor consequence relations: The flat case. *Journal of Applied Non-Classical Logics*, 12:63–84, 2002.

[33] Philippe Besnard and Torsten Schaub. Signed systems for paraconsistent reasoning. *Journal of Automated Reasoning*, 20:191–213, 1998.

[34] Kristof De Clercq. Two new strategies for inconsistency-adaptive logics. *Logic and Logical Philosophy*, 8:65–80, 2000. Appeared 2002.

[35] Jaakko Hintikka. *Inquiry as Inquiry: A Logic of Scientific Discovery*. Kluwer, Dordrecht, 1999.

[36] Larry Laudan. *Progress and its Problems*. University of California Press, Berkeley, 1977.

[37] Iddo Lev. Preferential systems for plausible non-classical reasoning. Master's thesis, Department of Computer Science, Tel-Aviv University, 2000. Unpublished M.A. dissertation.

[38] Joke Meheus. Adaptive logic in scientific discovery: the case of Clausius. *Logique et Analyse*, 143–144:359–389, 1993. Appeared 1996.

[39] Joke Meheus. *Wetenschappelijke ontdekking en creativiteit. Een poging tot theorievorming op basis van een conceptuele, methodologische en logische studie*. PhD thesis, Universiteit Gent (Belgium), 1997. Unpublished PhD thesis.

[40] Joke Meheus. Deductive and ampliative adaptive logics as tools in the study of creativity. *Foundations of Science*, 4:325–336, 1999.

[41] Joke Meheus. Model-based reasoning in creative processes. In Lorenzo Magnani, Nancy Nersessian, and Paul Thagard, editors, *Model-Based Reasoning in Scientific Discovery*, pages 199–217. Kluwer/Plenum, Dordrecht, 1999.

[42] Joke Meheus. The positivists' approach to scientific discovery. *Philosophica*, 64:81–108, 1999. Appeared 2001.

[43] Joke Meheus. Analogical reasoning in creative problem solving processes: Logico-philosophical perspectives. In Fernand Hallyn, editor, *Metaphor and Analogy in the Sciences*, pages 17–34. Kluwer, Dordrecht, 2000.

[44] Joke Meheus. An extremely rich paraconsistent logic and the adaptive logic based on it. In Batens et al. [29], pages 189–201.

[45] Joke Meheus. Adaptive logics for question evocation. *Logique et Analyse*, 173–175:135–164, 2001. Appeared 2003.

[46] Joke Meheus. An adaptive logic based on Jaśkowski's approach to paraconsistency. *Journal of Philosophical Logic*, in print.

[47] Joke Meheus and Diderik Batens. Steering problem solving between cliff incoherence and cliff solitude. *Philosophica*, 58:153–187, 1996. Appeared 1998.

[48] Joke Meheus and Dagmar Provijn. Direct dynamic proofs for classical compatibility. *Logique et Analyse*, 185–188:305–317, 2004. Appeared 2005.

[49] Marek Nasieniewski. *Logiki adaptujace sprzeczność (Inconsistency adapting logics)*. Phd thesis, Chair of Logic, N. Copernicus University, Toruń, Poland, 2002.

[50] Marek Nasieniewski. An adaptive logic based on Jaśkowski's logic $\mathbf{D}_2$. *Logique et Analyse*, 185–188:287–304, 2004. Appeared 2005.

[51] Thomas Nickles. What is a problem that we may solve it? *Synthese*, 47:85–118, 1981.

[52] Thomas Nickles. Reconstructing science: discovery and experiment. In Diderik Batens and Jean Paul Van Bendegem, editors, *Theory and Experiment. Recent Insights and New Perspectives on Their Relation*, pages 33–54. Reidel, Dordrecht, 1988.

[53] Thomas Nickles. Discovery logics. *Philosophica*, 45:7–32, 1990.

[54] Thomas Nickles. Can scientific constraints be violated rationally? In *Scientific Discovery, Logic, and Rationality* [56], pages 285–315.

[55] Thomas Nickles. Scientific discovery and the future of philosophy of science. In *Scientific Discovery, Logic, and Rationality* [56], pages 1–59.

[56] Thomas Nickles, editor. *Scientific Discovery, Logic, and Rationality*. Reidel, Dordrecht, 1980.

[57] Thomas Nickles, editor. *Scientific Discovery: Case Studies*. Reidel, Dordrecht, 1980.

[58] Graham Priest. *In Contradiction. A Study of the Transconsistent*. Nijhoff, Dordrecht, 1987.

[59] Graham Priest. Minimally inconsistent **LP**. *Studia Logica*, 50:321–331, 1991.

[60] Graham Priest. Paraconsistent logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2. Kluwer, Dordrecht, 2 edition, in print.

[61] Kurt Schütte. *Beweistheorie*. Springer, Berlin, 1960.

[62] Dudley Shapere. Logic and the philosophical interpretation of science. In Paul Weingartner, editor, *Alternative Logics. Do sciences need them?*, pages 41–54. Springer, Berlin, Heidelberg, 2004.

[63] Raymond M. Smullyan. *First Order Logic*. Dover, New York, 1995. Original edition: Springer, 1968.

[64] Liza Verhoeven. Proof theories for some prioritized consequence relations. *Logique et Analyse*, 183–184:325–344, 2003. appeared 2005.

[65] Andrzej Wiśniewski. *The Posing of Questions. Logical Foundations of Erotetic Inferences*. Kluwer, Dordrecht, 1995.

[66] Andrzej Wiśniewski. The logic of questions as a theory of erotetic arguments. *Synthese*, 109:1–25, 1996.

[67] Andrzej Wiśniewski. Questions and inferences. *Logique at Analyse*, 173–175:5–43, 2001.

[68] Andrzej Wiśniewski. Weak epistemic logics, immediate consequence, and paraconsistency. *Logique et Analyse*, 177–178:175–184, 2002.

3

CASE STUDIES ON ABDUCTION

How can the example of Kepler's great scientific discovery illuminate the problematical nature of abduction?

Ana Paula Rocha C. F. Silva
Centro de Filosofia das Ciências da Universidade de Lisboa
anapfsilva@netcabo.pt

> "Kepler was forever trying experiments with his figures. No bad luck, not dozens of negative results, which other men reckon failures, could discourage him from trying again…Each step was made deliberately, and for sound reasons; and few of Kepler's "failures" failed to throw some light on the problems he had in hand." (Peirce, C.S.)[1]

Introduction

Suppose Charles Sanders Peirce was alive today. Certainly he would see some of his ideas being rejected as energetically as they were several decades ago. We know that one of the reasons for the pioneering character of Peirce's work concerns the problematical nature of his concept of abduction, a concept with which Peirce aims to explain what has never been explained – the process of scientific discovery, *one by which* "theories and conceptions are engendered" (CP 5.590). At least in this point Peirce was - and still is - far ahead of his time[2]. The same happened three hundred years before with another great mathematician, astronomer and thinker: Johannes Kepler! In fact, Kepler had neither predecessors[3] nor contemporaries[4] in his

1 Peirce, C.S. (1892).

2 Hanson, N.R. (1965), pp. 42-43

3 Cohen, J.B. (1985) p.128

4 « Kepler… est, pourrait-on dire, à la fois en avance, et en retard sur ses contemporaines » - Koyré, A. (1961). p. 119

program of reforming astronomical theory, of explaining celestial physics through the search of the natural causes of planetary motions[5].

As we know, Peirce took Kepler as a major example of the abduction process, "the *eternal exemplar* of scientific reasoning". (CP 2.96) As Peirce writes repeatedly, Kepler's hypothesis is "the greatest piece of Retroductive Reasoning ever performed" (CP 1.74).

Now, this paper considers the following correlated questions:

- Why was Peirce so much interested in Kepler's example?
- Why did *Peirce* choose a discovery with 300 years old as the best, *eternal example,* of his concept of abduction?
- How can the example of Kepler's great discovery illuminate the problematical, almost paradoxical, nature of abduction?

In order to look for a tentative answer to those questions, this paper will develop a double movement based on the analysis of some texts from Peirce and some from Kepler. In fact, it is a balanced double movement since I will consider texts from Peirce about Kepler[6] and, inversely, texts from Kepler, mainly from *Astronomia Nova* (1609), which can be at the root of the importance *Peirce awards to* Kepler's discovery. The aim will be to look for the main phases of the processes of discovery which Kepler has passed through, and

5 Kepler, himself, in *Astronomia Nova* (1609) proclaims the revolutionary character of his astronomy. For further developments, see Cohen, J. B. (1985) p.128

6 Concerning Peirce, the most important references to Kepler can be found in several moments and in several texts:

- In 1877 - The celebrated *The Fixation of Belief* * (CP 5.375)
- In 1877 - *The Theory of Probable Inference* (Writings of C.S.Peirce – Vol. 4)
- In 1892 – a full article dedicated to Kepler and precisely untitled: *Kepler* (1892)
- In 1896 – *Kinds of Reasoning* (CP 1.65-1.74)
- In 1901 – *Abduction* (CP 7.218-7.220)
- In 1902 – *Proposed Work in Logic* (CP 2.96), perhaps the one which offers the most accurate presentation of Kepler's discovery.

*As Peirce recognizes, it was only in 1892, at the occasion of a series of lectures on History of Science from Copernicus to Newton, that he felt qualified to talk seriously about the discovery process which led Kepler to the results later named as Kepler´s first and second law. In fact, in 1893, Peirce wrote that he was ashamed to confess that the text *The Fixation of Belief* contained a foolish remark about Kepler. He even says that, in 1877, when he wrote *The Fixation of Belief,* he had never truly studied the original book from Kepler

which allowed him to arrive to the hypothetical conjecture of Mars' elliptical orbit.

Astronomia Nova (AN) is a strange work, very different from what we consider nowadays as a scientific book. The book contains an historical presentation and extremely detailed description of the discovery process which led Kepler to the elliptic orbit of planets. Kepler believed that his readers would wish to follow "the roads by which men arrive at their insights into celestial matters."[7] In the introduction of AN, he compares his work of discovery with those of a Columbus and a Magellan in whose narratives people find great entertainment. That is why AN looks like a diary in which, day after day, Kepler remarks the results he obtains and presents the series of false starts and failures he encounters on his work[8]. All mixed with religious-philosophical considerations and autobiographical details[9]. However, according to Peirce, we are face to face to a work which demands "the most vigorous exercise of all the powers of reasoning from beginning to end". (CP 1.71)

In fact, in all the History of Science, Kepler is an extremely valuable case of self-reflection on his own scientific discoveries[10]. In AN he fully describes how he generated multiple, temporary, and substituting hypotheses. Kepler's great discovery — the orbits of the planets are elliptical rather than circular — was initially based on an anomalous observation: the longitudes of Mars did not fit circular orbits[11]. As Kepler himself tells us, before even dreaming

7 see Koestler, A . (1963), p 261

8 see Voelkel, J.R. (2001), who emphasises mostly this point

9 Kozhamthadam, J. (1994) studies precisely this issue

10 William H. Donahue, among others, has shown that the account Kepler offers his readers is not a true history of his research-something Kepler never claimed-but rather a didactic or rhetorical pseudohistory. But the question of why Kepler chooses this form of exposition has not been addressed. Voelkel tries to find the answer to this question in the context of the composition of the *Astronomia Nova* and in Kepler's relation to the contemporary astronomical community. He argues that the conceptual and stylistic features of the *Astronomia Nova* are intimately related: Kepler purposely choose this form of exposition precisely because of the response he knew to expect from the astronomical community to the revolutionary changes in astronomical methodology he was proposing. (Voelkel, 2001)

11 "…if I had thought I could ignore eight minutes of longitude, in bisecting the eccentricity I would already have made enough of a correction in the hypothesis found in ch. 16. Now, because they could not have been ignored, these eight minutes alone will have led

that, instead of circles, ellipses were the best explanation, he tried several other forms of explanation.

Why was Peirce so much interested in Kepler's example?

Now, let us came to our main point. Why was Peirce so much interested in Kepler's example?

I will point out nine reasons:

1) According to Peirce, Kepler's discoveries of the paths of planets and the law of their varying speeds are **one of the most important scientific achievements of all times** "an achievement by far the most triumphant unravelment of facts ever performed". (*Kepler*,p.250). A discovery which Peirce considers to be in a higher position than the deciphering of the cuneiform inscriptions or the hieroglyphics. (*Kepler*, p.250)

2) According to Peirce, Kepler was the first astronomer to be **interested in the causes of the phenomena**[12]. When Newton, for instance, used the word hypothesis in the famous expression *Hypotheses non fingo*, he intended to give a general formula for the motions of the heavenly bodies, but he "was not undertaking to mount the causes" (CP 2.707). On the contrary, Kepler proposed not only to give a more rigorous kinematics to the motion of planets, but also to individuate the cause of the motion itself.[13] In Peirce's words, "Kepler, looking at the matter dynamically, thought it must have something to do with causing the planets to move in their orbits". (CP 1:72)

---

the way to the reformation of all of astronomy, and have constituted the material for a great part of the present work." – Kepler in AN, p.286

12 Holton (1973) wrote: "Kepler's genius lies in his early search for a physics of the solar system. He is the first to look for a universal physical law based on terrestrial mechanics to comprehend the whole universe in its quantitative details" (p. 71) .

13 "I am much occupied with the investigation of the physical causes. My aim in this is to show that the celestial machine is to be likened not to a divine organism but rather to a clockwork . . . , insofar as nearly all the manifold movements are carried out by means of a single, quite simple magnetic force, . . . Moreover, I show how this physical conception is to be presented through calculation and geometry." - Kepler, in a 1605 letter to von Hohenburg, cited in Holton, 1973, p. 72

3) Kepler performs Peirce's ideal of **scientific method**[14]. As Peirce write: "Kepler, looking at the matter **dynamically**, thought it must have something to do with causing the planets to move in their orbits. He is one of the few thinkers who have taken their readers fully into their confidence as to what their method really has been". (CP 6.604) Also, according to Peirce, Kepler "… saved the appearances as far as heliocentric longitudes went; and that would have satisfied many an astronomer. But Kepler could not be satisfied, since the theory did not agree with the latitudes or geocentric longitudes; and by far the greater part of his labour came after he had obtained this vicarious hypothesis". (*Kepler*, p.256)

4) According to Peirce, Kepler had "an admirable **method of thinking**"[15]. Peirce repeatedly praises the "perfection of Kepler's ratiocination", his "gigantic power of right reasoning" (*Kepler*, p.250, 252), his "keen logical sense". (CP 1.74)

5) According to Peirce, Kepler had the first quality required for high reasoning: **imagination**[16]. He said that "Kepler's fecund imagination strikes

14 For instance, Kepler said in the introduction to the AN: "Now my first step in investigating the physical causes of the motions was to demonstrate that [the planes of] all the eccentrics intersect in no other place than the very centre of the solar body (not some nearby point), contrary to what Copernicus and Brahe thought…I have to look again and again at what I am doing, so as to avoid setting up a new method.. I have demonstrated in the first part of the work that exactly the same things can result or be presented by this new method as are presented by their ancient method… However, my method is in agreement with physical causes, and their old one is in disagreement" (AN p. 48,49)

15 Peirce continues: "All the endowments of Kepler's intellect and heart seem to have been concentrated upon one function, that of reasoning. In his great work on Mars, he has laid bare to us all the operations of his mind during the whole research; and what better sign of the perfection of his ratiocination could there be than that no better pathway could be found by which to lead another thought to the same conclusion than that his own had broken in the first instance. His admirable method of thinking consisted in forming in his mind a diagrammatic or outline representation of the entangled state of things before him, omitting all that was accidental, observing suggestive relations between the parts of his diagram, performing divers experiments upon it, or upon the natural objects, and noting the results. (*Kepler*, p.254, 255).

16 Also Holton (1973) wrote about this: "[Modern scientists are] taught to hide behind a rigorous structure the actual steps of discovery-those guesses, errors, and occasional strokes of good luck without which creative scientific work does not usually occur. But Kepler's embarrassing candor and intense emotional involvement force him to give us a detailed account of his tortuous process . . . . He gives us lengthy accounts of his failures,

every reader" and further, he stresses that "Kepler's imagination makes the clothing and the flesh drop off, and the apparition of the naked skeleton of truth to stand revealed before him". (*Kepler,* p.255)

6) According to Peirce, Kepler was able to recognize his own **errors**. "No man was ever more coolly sensible of his own faults and weakness" than Kepler. (*Kepler,* p.255) Moreover, according to Peirce, Kepler was also able to overcome those errors: "No bad luck, not dozens of negative results, which other men reckon failures, could discourage him from trying again". *(Kepler,* p.256)

7) According to Peirce, Kepler was endowed with a marvelous capacity of **guessing**, a capacity which gave him "an inward power, not sufficient to reach the truth by itself, but yet supplying an essential factor to the influences carrying their minds to the truth". (CP 1.80) This capacity or *lume naturale* was a powerful guide in the scientific discovery: "Kepler did not conclude from this that the orbit really was an ellipse; but it did incline him to that idea so much".[17] (CP 2.96)

8) According to Peirce, Kepler had not only the intellectual qualities necessary to scientific discovery, but also the **moral qualities required by scientific character**: "an incomparable **e**nergy and courage", (CP 5.363) "an admirable curiosity about all the ways of nature", (*Kepler*,p.252) a "great intellectual labour." (CP 1.72) Peirce characterized the moral fibre of Kepler's

though sometimes they are tinged with ill-concealed pride in the difficulty of his task. With rich imagination he frequently finds analogies from every phase of life, exalted or commonplace. He is apt to interrupt his scientific thoughts, either with exhortations to the reader to follow a little longer through the almost unreadable account, or with trivial side issues and textual quibbling, or with personal anecdotes or delighted exclamations about some new geometrical relation, a numerological or musical analogy." (pp. 69-70)

17 "at a certain stage of his eternal exemplar of scientific reasoning, he found that the observed longitudes of Mars, were (within the possible limits of error of the observations) such as they would be if Mars moved in an ellipse. The facts were thus, in so far, a *likeness* of those of motion in an elliptic orbit. Kepler did not conclude from this that the orbit really was an ellipse; but it did incline him to that idea so much as to decide him to undertake to ascertain whether virtual predictions about the latitudes and parallaxes based on this hypothesis would be verified or not. This probational adoption of the hypothesis was an Abduction…"(CP 2.96)

scientific character in the following words: "Kepler ran down the truth like an indefatigable detective, with hardly a wasted day." (Peirce, MS 514) and also "… all the endowments of Kepler's intellect and heart seem to have been concentrated upon the function of reasoning." (*Kepler*, p.255)

9) Finally, according to Peirce, Kepler illustrates the most ingenious use of **abductive reasoning** in the history of science. But, perhaps, may be important to remember what Peirce wrote: "At each stage of his long investigation, Kepler has a theory which is approximately true, since it approximately satisfies the observations (that is, within 8', which is less than any but Tycho's observations could decisively pronounce an error), and he proceeds to modify this theory, after the most careful and judicious reflection, in such a way as to render it more rational or closer to the observed fact…. Thus, never modifying his theory capriciously, but always with a sound and rational motive for just the modification he makes, it follows that when he finally reaches a modification - of most striking simplicity and rationality – which exactly satisfies the observations, it stands upon a totally different logical footing from what it would if it had been struck out at random, or the reader knows not how, and had been found to satisfy the observation. Kepler shows his keen logical sense in detailing the whole process by which he finally arrived at the true orbit. This is the greatest piece of Retroductive reasoning ever performed." (CP 1.71-74)

Five hierarchical levels within the concept of abduction.

Let us came to our main question: Why is the example of Kepler's discovery "the greatest piece of Retroductive Reasoning ever performed"? How can this example illuminate the problematical nature of abduction?

My hypothesis is that we ought to consider several hierarchical levels within the concept of abduction. May be we succeed to get a description able to clarify the creative processes of Kepler´s scientific discovery.

At a lower level, abduction would consist of finding the correct solution by the progressive elimination of mistaken hypothesis within of a limited

number of possibilities. At this first level, abduction would be restricted to a process of testing and error.

At a slightly higher level, abduction would allow us to go directly to the good solution within a limited number of possibilities. But if we find the good solution by chance, then we are at a minor intermediate level (second level) than if we approach the truth by guessing (third level)

Now, if we incline ourselves to the good solution, not within a limited number of possibilities but face to an unlimited number of hypotheses, then we would need a huge abductive power. And, in the case of approaching the truth within an unlimited number of possibilities, yet we can consider two levels: the fist and lower, if those unlimited possibilities are all in a same context (fourth level); the second and higher, if those possibilities belong to various contexts (fifth level).

So we have:

- Limited Number of Hypothesis
  - Progressive Elimination of Mistaken Hypothesis — Trial and Error
  - Direct Finding
    - Chance (2nd level)
    - Guess (3rd level)
- Unlimited Number of Hypothesis
  - Possibilities are all in a same Context (4th level)
  - Possibilities belong to various Conext (5th level)

Let us now look carefully to the Kepler's description of his own discoveries, those which are at the basis to the hypothetical conjecture of Mars' elliptical orbit. May be we succeed to get a better understanding of Peirce's interest.

First Point: *Abduction as a progressive elimination of mistaken hypothesis, a process of trial and error* (first level)

AN is rich in descriptions of Kepler's errors and mistakes. Instead of presenting the positive results of his efforts, Kepler explicitly intends to

"display these occasions [errors[18] and meanderings] scrupulously" [19] (AN p. 95). And, in fact, further that Kepler repeated claim[20], error played an important role in some of his scientific discoveries. That is the case of Kepler's rejection of the circular orbit and of his subsequent acceptance of the oval, and afterwards of the elliptical, shape of the planetary orbit.

In the summaries of AN chapter 44, Kepler wrote that it was demonstrated, by two arguments, that the planet's orbit was not a circle, but an oval figure. Kepler said about the first argument:

"… those things demonstrated in chapters 41 and 42 are presupposed. Now a perfect circle results in certain distances, and the diameter of this circle was found in chapter 42. But the observations which were reintroduced in chapter 41 require different distances, namely, ones that are shorter at the sides. But an oval figure admits such distances. Therefore, the orbit is an oval.[21] (AN p. 95)

If we put these words in the formal way of Peirce's abduction, we get:

"The surprising fact, C, is observed;

But if A [an explanatory hypothesis] were true, C would be a matter of course,

Hence, there is reason to suspect that A is true" (CP 5.189)

So, we have:

18 See Note 1 in A.N. p.79 where Donahue wrote: "The reader must beaware that Kepler is presenting a historical account of his enquiry…. What is remarkable, however, is his sense that mistakes are important, that we approach truth by first being wrong."

19 Kepler adds: "we do not simply ignore the errors" (AN p.78), "something among those things we have assumed must be false" (AN p.283), "If I had embarked upon this path a little more thoughtfully, I might have immediately arrived at the truth of the matter. But since I was blind from desire [to explain the deviation from a circular orbit] I did not pay attention to each and every part . . . and thus entered into new labyrinths". (AN p.455,456) And when mentioning his "wearisome method", Kepler confesses he had gone through it at least seventy times." (AN p.256).

20 For instance, Kepler wrote: "My first error was to suppose that the path of the planet is a perfect circle, a supposition that was all the more noxious a thief of time the more it was endowed with the authority of all philosophers, and the more convenient it was for metaphysics in particular" (AN p.417)

21 We think that the second argument presented by Kepler is also in an abductive form: "In the second argument, the same things are presupposed as in chapter 43. The elapsed times shown to us by experience are not admitted by a circular figure, but are admitted by an oval. Therefore, the orbit of the planet is an oval." (AN p. 95)

1) The surprising fact, C *"The observations which were reintroduced in chapter 41 require different distances, namely, ones that are shorter at the sides"*, is observed;
2) But if A "*the orbit is an oval figure - admits such distances"* were true, C *"The observations which were reintroduced in chapter 41 require different distances, namely, ones that are shorter at the sides",* would be a matter of course,
3) Hence, there is reason to suspect that *A the orbit is an oval"* is true.

What we have here is a first level abduction. What we have here is the abandon of a **mistaken** hypothesis (the hypothesis about the circular orbit) because the observation of different distances did not fit to a circle. Face to the awareness of such "errors", Kepler's solution is to find another type of curve (the oval) able to support the new data. We know that, afterwards, Kepler rejected the oval orbit and accepted the ellipse as the shape of the planetary orbit. So, we can almost say that abduction is here restricted to a process of progressive elimination of mistaken hypothesis, a process, we could say, of **trial and error**. Of course this process needs induction in order to prove to be false the hypothesis which are to be eliminated.

Yet, we must stress that this abductive procedure may only have a sense in a limited number of possibilities. Because the orbit should be a curve, it could only be circular, oval or elliptic. In other words, face to the three possibilities for the path of a planet, Kepler would have eliminated, first, the circular path, second, the oval, and finally he would got the ellipse.

That's why, when he rejected the oval in favour of the ellipse he reacted emotionally: "O ridiculous me! To think that the reciprocation on the diameter could not be the way to the ellipse!…no figure is left for the planet to follow other than a perfectly elliptical one." (AN, p.576)[22]. Now, we under-

22 "I could not discovery why the planet,… would rather follow as elliptical path, as shown by the equations. O ridicolous me! To think that the reciprocation on the diameter could not be the way to the ellipse! So it came to me as no small revelation that through the reciprocation an ellipse was generated. This will be made clear in the following chapter, where it will be demonstrated at the same time, through the agreement of arguments from physical principles with the body of experience, mentioned in this chapter, that is contained in the observations and in the vicarious hypothesis, that no figure is left for the planet to follow other than a perfectly elliptical one." (AN, p.576)

stand why, as pointed earlier, Peirce praises so much Kepler's capacity of recognizing and overcoming his own errors[23].

However, what we would like to stress is that only at a very elementary level can abduction be reduced to a process of trial and error. Only at a very elementary level can abduction be thought out as a process of finding the truth by the progressive elimination, by induction, of mistaken and wrong hypothesis.

Second Point: *Abduction as a way of getting directly to good solution by chance* (second level)

In other texts, it is not error, but chance[24] which Kepler emphasises as a fundamental element of his scientific endeavour. For instance, it was by chance that Kepler visited Tycho when his disciple, Christian, had in hands the theory of Mars. As Kepler says: "Had Christian been treating a different planet, I would have started on it as well". (AN p. 185)[25].

In other passages, Kepler is even clearer. For instance, when he says: "It is a happy accident that the distances found in chapter 51 also inform us about this" (AN p.526). We are now face to an abductive approach which does not demand, any longer, any indirect process of trial and error. Abduction is now a process which allows going directly to the solution; which allows jumping preliminary procedures. And that capacity appears as to be caused by chance.

However, if it was just this capacity of dealing with chance (or with error), that would be at the root of Peirce's interest in Kepler's "eternal exam-

23 "If I had embarked upon this path a little more thoughtfully, I might have immediately arrived at the truth of the matter . But since I was blind from desire [to explain the deviation from a circular orbit] I did not pay attention to each and every part . . . and thus entered into new labyrinths, from which we will have to extract ourselves." (AN pp. 455, 456)

24 In the introduction to the Astronomia Nova, Kepler claimed: "Here it is a question not only of leading the reader to an understanding of the subject matter in the easiest way, but also, chiefly, of the arguments, meanderings, or even chance occurrences by which I the author first came upon that understanding." (AN p.78)

25 "started on it as well …I therefore once again think it to have happened by divine arrangement, that I arrived at the same time in which he was intent upon Mars, whose motions provide the only possible access to the hidden secrets of astronomy, without which we would remain forever ignorant of those secrets." (AN p. 185)

ple", then, we could almost say that, by choosing Kepler, Peirce would have chosen **bad or poor examples** of abduction.

In fact, the logical aim of abduction would have been entirely absorbed by an aleatory, casual element. In both cases (trial and error and chance) we would face poor ways of thinking the nature of abductive reasoning. **But, this is not what really happens.**

Third point: *Abduction as the process of going directly to the good solution by guessing* (third level)

In many other instances, Kepler refers the capacity of guessing which underlies his scientific discoveries: "Now since I suspected what proved to be true, that the hypothesis was inadequate, I entered upon the work girded with the preconceived opinions expressed in my *Mysterium Cosmographicum*. (AN p. 184)[26] or:" So it came to me as no small revelation that through the reciprocation an ellipse was generated" (AN p.576).

Sometimes, Kepler himself is very much astonished with his own capacity of guessing (to use the words from Peirce) or in Kepler's words: "A guess[27] as to how small this discrepancy is going to turn out to be" (AN p.468). He even claims to have been inspired by a spirit in his search for the truth: "in order either to confirm or undermine this speculation by observations" (AN p.309), "a sprite whispered to me that this illusion arises thus: Copernicus's annual orb[28] or Ptolemy's epicycle, is not everywhere distant from that centre

26 "…From my Mysterium Cosmographicum – that an equant might be introduced into the theory of the sun (or, as I call it, the theory of the Ptolemaic epycicle" (AN, p.306) or "Also, for the best reasons, one of our suppositions" (AN, p.548) or "…that suspicion of mine arising…" (AN p.306)

27 Kiikeri (1999) defends "that inference to the best explanation (IBE)-account could show why Kepler´s discovery was not merely a guess (as Newton famously claimed) but a justified result which was based on a coherent world view, strict methodological standards, and sound reasoning"

28 Copernicus in his picture of the universe had placed the sun in the centre. However he had, in presenting the planet theories, assumed as centre of the universe, not the sun itself but rather the centre of the earth's orbit, which was somewhat to the side of the sun, and referred all his calculations to this. Tycho Brahe, in his system, had made an assumption corresponding to this. Both of them, on the strength of these assumptions, erected their planet theories on oppositions to the so-called mean sun instead of to the true one. Kepler required that all values be referred to the true sun (Caspar, (1993) p.128)

about which by supposition it is sweeping out equal angles in equal times." (AN p.306). That is to say: in these instances, Kepler's discovery would have been choosing as an example of a third level abduction.

Fourth point: *abduction as a process of reaching a solution within an unlimited number of other sorts of hypothesis* (fourth level)

We know that the Greeks assumed that the celestial bodies moved uniformly in perfect circles or in a combination of uniform circular motions. But uniform motion around circles did not account for the complicated movements observed, and the Greeks faced the problem of "saving the phenomena" through the mathematical model based on the theory of eccentrics and epicycles. Those theories, with their many modifications, constituted the main mathematics scheme for all astronomical calculations during about eighteen centuries.

We also know that the Greeks were basically oriented towards the description of motions and the finding of mathematical regularities. They were not committed to find the causes of the motion of celestial bodies, they renounced the investigation of the true nature of the physical system of the world. So, their attention was mainly given to the kinematics and not to the dynamics of celestial bodies.

On the contrary, Kepler proposed not only to "save the phenomena, the appearances"[29], and to give a more rigorous kinematics to the motion of planets, but also to individuate the cause of the motion itself[30].

It is here that we arrive to the referred fourth level: *abduction as a process of finding one solution within an unlimited number of other sorts of hypothesis* in a same context.

Kepler is the first to look for a universal physical law based on terrestrial mechanics to comprehend the whole universe in its quantitative details. As

29 "I consequently showed, using the arguments presented already in Part I, that an eccentric can be false, yet answer for the appearances within five minutes or better, provided that the centre of the equant be correctly known" (AN pp. 185, 186)

30 As Caspar (1993) put it: "It is Kepler's greatest service that he substituted a dynamic system for the formal schemes of the earlier astronomers,the law of nature for mathematical rule, and causal explanation for the mathematical description of motion" (p . 136).

he said: "I treat all of astronomy by means of physical causes[31] rather than fictitious hypotheses" . Of course, as Kepler explicitly recognizes, the rise of truly innovative hypotheses may only occur if one opens one's mind to that possibility. That is precisely what Kepler tells us: "When I was giving the physical cause of the Ptolemaic *equant* or of the Copernican-Tychonic second epicycle, I raised an objection against myself …I decided *to leave that speculation open,* until the matter were clearer to astronomers I nevertheless entertained a suspicion…" (AN p.305)[32]

Within such openness, abduction faces, not a limited number of already known possibilities (as in the case of the circle, the oval and the ellipse) but, on the contrary, an unlimited number of other sorts of hypothesis. As Kepler wrote: "At the beginning there was great controversy between us as to whether it were possible to set up **another sort of hypothesis** which would express to a hair's breadth so many positions of the planet, and whether it were possible for the former hypothesis to be false despite its having accomplished this so far over the entire circuit of the zodiac". (AN p.184)[33]. Here, Kepler is already out of the effective context of his time.

Fifth point: *abduction as a process of reaching a solution within an unlimited number of other sorts of hypothesis belong to various context* (fifth level)

We are now facing a fifth level: *abduction as a process of finding one solution within an unlimited number of other sorts of hypothesis.*

In fact, for Kepler the solutions found in such an open field appears as complete novelty, an unprecedented issue, almost an absurdity. That is why, taking in consideration the problems which the several systems of Ptolemy,

31 Kepler wrote at the beginning of his AN that "the testimony of the ages confirmed that the motions of the planets are orbicular. But that is an immediate presumption of reason, reflected in experience, that their gyrations are perfect circles. However when experience teaches something deviate from a simple circular path, it gives rise to a powerful sense of wonder, which at length drives men to look into causes."

32 or "when I saw that as the physical causes introduced in chapter 45 were the more skilfully,…, I greatly congratulated myself,…On the other hand, ..I did not rest until I add established a more certain and direct way, and at the same time I began to suspect…" (AN p.488)

33 Besides Kepler said in chapter 53: "Since we are establishing new hypotheses here" (AN p.529).

Copernicus and Tycho Brahe put in relation to the point *equanto*, to the law of the areas and to the physical causes of the planetary movements[34], Kepler asks: "What, I ask, is this **novelty unprecedented** in astronomy, **this unlikely absurdity**?" (AN p.306)[35]

What we would like to stress is that we are reaching here a deeper concept of abduction. Abduction is no longer the progressive elimination of error. It is no longer a process witch could be bestowed by chance. Nor it can be explained, only by the guessing strength of the *lume naturale*[36]. Besides, abduction doesn't just involve a process of finding one solution within an unlimited number of other sorts of hypothesis in a same context. Abduction demands now the courage, the "admirable curiosity" (*Kepler*, p.252), the "incomparable energy" (CP 5.363), the moral fibre of Kepler's scientific character, which enables him to putt an almost absurd hypothesis, to leave aside the possible known contexts. When Kepler said: "And this is the end

34 Although Copernicus's work has been regarding as significant for putting forward the idea that the earth travels around the sun, sixteenth-century astronomers largely ignored that claim, which violated Aristotelian physics and apparently contradicted the testimony of Holy Scripture. Instead, they were attracted by Copernicus's novel form of mathematical planetary theory, which eliminated Ptolemy's equant, a mechanism that caused the centre of a planet's epicycle to travel no uniformly around its eccentric and thus violated the precept that planetary theories should be composed of compounds of uniform circular motion. (Voelkel, 2001)

35 "…Wouldn't it seem more worthy of belief that the sun (for Copernicus) or the centre of the planetary system (Tychonic) or the body of the planet (for Ptolemy) would in certain places be farther from, and in others nearer to, the selected point of uniform motion (at rest for Copernicus and Tycho, and moving round on the circumference of the eccentric for Ptolemy) – and this especially on the line of apsides? And for this, that suspicion of mine arising from my Mysterium cosmographicum - that an equant might be introduced into the theory of the sun (or, as I call it, the theory of the Ptolemaic epicycle) - provides a convenient occasion?" (AN p.306)

36 "In examining the reasoning of those physicists who gave to modern science the initial propulsion which has insured its healthful life ever since, we are struck with the great, though not absolutely decisive, weight they allowed to instinctive judgments. Galileo appeals to il lume naturale at the most critical stages of his reasoning. Kepler, Gilbert, and Harvey — not to speak of Copernicus — substantially rely upon an inward power, not sufficient to reach the truth by itself, but yet supplying an essential factor to the influences carrying their minds to the truth. It is certain that the only hope of retroductive reasoning ever reaching the truth is that there may be some natural tendency toward an agreement between the ideas which suggest themselves to the human mind and those which are concerned in the laws of nature." (CP 1:80)

of the second part, in which I have imitated the ancients…Now, following my method, I begin the whole inquiry anew",[37] Kepler was jumping outside the restricted contexts of his time, out of the effective paradigms, out of all and any authority[38]. His work offers an example of conceptual change in cosmologic and astronomic theory.

Therefore, we must not be surprised with his words:

"But then, by two arguments, which I have already presented…, I concluded that the breadth of the lunule is to be taken as only half that, namely, 429, or more correctly, 432,[39] ... I therefore began to think of the causes and the manner by which a lunule of such a breadth might be cut off. While I was anxiously turning this thought over in my mind, reflecting that absolutely nothing was accomplished by chapter 45 and consequently my triumph over Mars was futile, quite by chance[40], I hit upon the secant of the angle 5°18', which is the measure of the greatest optical equation. And when I saw that this was 100,429, it was as if I were awakened from sleep to see a new light, and I began to reason thus." (AN 543).

Flying without a net, alone in front of the nude truth, exposed to a new light, Kepler is like waking up from a dream. From now on, Kepler has no reasons to feel that "vague and indefinite fear that one experiences when first confronted by an unknown enemy"(AN p. 79)[41].

37 "… the circumstances are recounted that led me to suspect the operation of an equant circle in the theory of the sun" (AN p. 86, 87),

38 In fact, Kepler said in the introduction to the AN that his first step in his investigation was to demonstrate something contrary to what Copernicus and Brahe thought: "…I have to look again and again at what I am doing, so as to avoid setting up a new method." And like he own said he had demonstrated in the first part of his work that exactly the same things can result or be presented by his new method as were presented by their ancient method. But the great difference was that Kepler'method was in agreement with physical causes, and their old one was in disagreement. (AN p.48,49)

39 Like Peirce said "the reason for the most essential step of Kepler's great discovery was due to his remarking that two numbers, one 429, the other 432, which seemed to have no connection with one another were nearly equal." (*Kepler,* p.256)

40 it would be interesting investigate in which way Kepler mistakes the reasons of his own discovery.

41 Kepler has reasons to say "that which had tormented us for a long time in chapter 39 now surrenders to us in the statement of the truth we have perceived" (AN p.544).

He is enough brave and audacious to face (quote Peirce) "the apparition of the naked skeleton of truth" (*Kepler*, p.255).

And, in this case, we have to agree with Peirce and to consider that Kepler is an excellent example of abduction.

References

Aiton, E. J. (1969) «Kepler's Second Law of Planetary Motion», *Isis,* **60**, 75-90

Aiton, E. J. (1978) «Kepler's path to the construction and rejection of his first oval orbit for Mars», *Ann. of Science,* 35 (2), 173-190.

Aiton, E. J. (1975) « How Kepler discovered the elliptical orbit», *Math. Gaz.* 59 (410), 250-260.

Boccaletti, D. (2001) *From the Epycicles of the Greeksto Kepler's Ellipse – The Breakdown of the circle Paradigm* - <http://arxiv.org/abs/physics/0107009>

Caspar, M. (1993) *Kepler*, trans.C.Doris Hellman, New York: Dover Publications.

Cohen, J. B. (1985) «Kepler, Gilbert, and Galileo: A Revolution in the Physical Sciences?», in *Revolution In Science,* 126-145

Donahue, W. H. (1993) «Kepler's Transition to the Oval», *Journal for the History of Astronomy,* xxiv.

Gentner, D. et Al. (1997) «Analogical reasoning and conceptual change : A case study of Johannes Kepler», (Special issue) in *Journal of the Learning Sciences*, 6, 3-39

Gingerich, O. (1981) «Kepler», in the *Dictionary of Scientific Biography*, New York: vol. 7 and 8, p. 289-312.

Hanson, N. R. (1958) «The Logic of Discovery», *The Journal of Philosophy*, 55, 1073-1089.

Hanson, N. R. (1960) «More on The Logic of Discovery», *The Journal of Philosophy*, 57, 182-188.

Hanson, N. R. (1961) «Retroductive Inference», in Bernard Baumrin (Ed.), *Philosophy of Science, The Delaware Seminar. Vol.1,* New York, London: Interscience Publishers, 21-37.

Hanson, N. R. (1961) «Is there a logic of scientific discovery», in Feigl, H. and Maxwell, G., (Eds.), *Current Issues in the Philosophy of Science*, New York: Holt, Rinehart and Winston, 20-35.

Hanson, N. R. (1965) «Notes toward a logic of discovery», in Bernstein, R. J. (Ed.), *Perspectives on Peirce*, (1965), New Haven, London: Yale Univ. Press, 42-65.

Holton, G. (1973) *Thematic origins of scientific though,* Cambridge, MA: Harvard University Press.

Holton, G. (1988) «Johannes Kepler's Universe: Its Physics and Metaphysics», in *Thematic origins of scientific thought: Kepler to Einstein,* Cambridge, MA: Harvard University Press.

Kepler, J. (1609) *Astronomia Nova,* trans. W.H. Donahue, 1992, Cambridge: Cambridge University Press

Kiikeri, M. (2001) Abduction, IBE, and the Discovery of Kepler´s Ellipse - http://www.valt.helsinki.fi/kfil/matti/kiikeri.pdf

Kleiner, S. A. (1983) «A New Look at Kepler and Abductive Argument», Studies in History and Philosophy of Science, 14, 279-313.

Koestler, A. (1960) *The Watershed - A Biography of Johannes Kepler*, New York: Anchor Press.

Koestler, A. (1963) *The sleepwalkers*, New York: Grosset & Dunlap

Koyré, A. (1957) *From the Closed World to the Infinite Universe,* New York: Harper and Brothers.

Koyré, A. (1961) «La Révolution Astronomique – Copernic, Kepler, Borelli», Hermann

Koyré, A. (1973) *The Astronomical Revolution : Copernicus, Kepler, Borelli*, New York: Dover

Kozhamthadam, J. (1994) *The Discovery of Kepler's Laws: the interaction of science, philosophy and religion,* London: University of Notre Dame Press.

Kuhn, T. S. (1957) *The Copernican revolution: Planetary astronomy in the development of Western thought,* Cambridge, MA: Harvard University Press .

Niiniluoto, I. (1999) «Defending abduction», *Philosophy of Science,* 66, S436-S451.

Paavola, S. (2004) «Abduction Through Grammar, Critic and Methodeutic», *Transactions of the Charles S. Peirce Society,* 40-2: 245-270.

Paavola, S. (2004) «Abduction as a Logic and Methodology of Discovery: The Importance of Strategies», *Foundations of Science,* **9**(3), 267-283. <http://www.helsinki.fi/science/commens/papers/abductionstrategies.html>

Peirce, C. S. (1931-1958) *Collected Papers*, 8 vols., C. Hartshorne, P. Weiss and A. Burks (Eds.), Cambridge, MA: Harvard University Press.

Peirce, C. S. (1877) «The Fixation of Belief», in J.Buchler (Ed), *Philosophical Writings of Peirce,* 1955, New York: Dover Publications.

Peirce, C. S. (1892) «Kepler», in Philip P Wiener (Ed.), *Selected Writings*, 1966, New York: Dover Publications.

Peirce, C. S. (1839-1914) «Textos de Charles Sanders Peirce» in Mota, O. Silveira (Eds.) (1975), *Semiótica e Filosofia*, Introdução, selecção e tradução de Octanny Silveira da Mota e Leonidas Hegenberg, 2ª Edição, São Paulo: Cultrix

Peirce, C. S. (1839-1914) «Escritos Coligidos/Charles Sanders Peirce» in Mora d'Oliveira, e Henrique dos Santos, L. (Eds). (1980), *Escritos Coligidos Charles Sanders Peirce e Gotlob Frege*, Selecção e Tradução de Armando Mora d'Oliveira, Sérgio Pomerangblum e L.H. dos Santos, S. Paulo: Abril Cultural, Colecção Os Pensadores.

Peirce, C. S. *Writings of C.S.Peirce : A Chronological Edition*, Volume I - (1857-1866), Volume II (1867-1871), Volume III (1872 -1878), Volume IV (1879-1884) Volume V (1884-1886). Edited by the Peirce Edition Project (Indiana University Press, Bloomington, Indiana, 1982, 1984, 1986, 1989, 1993).

Peirce, C. S. (1909) *MS 514* <http://www.jfsowa.com/peirce/ms514.htm>

Richter, A. (1995) *Der Begriff der Abduktion bei Charles Sanders Peirce*, Frankfurt am Main: Lang, 83-93

Stephenson, B. (1987 reprint 1994) *Kepler's physical astronomy*, Princeton, NJ : Princeton University Press.

Weiss, P. (1952) «The Logic of the Creative Process», in Philip P. Wiener and Frederic H. Young (Eds.), *Studies in the Philosophy of C.S. Peirce.* Cambridge, MA: Harvard University Press, 166-182.

Wilson, C. (1968) «Kepler's Derivation of the elliptical path», *Isis,* 59, 5-25

Wirth**,** U**.** (1999) *Abductive reasoning in Peirce´s and Davidson´s account of interpretation*, in: Transactions of the Charles Sanders Peirce Society, XXXV (1999), S.115-128.

Wirth, U. (1998) *What is Abductive Inference*?, <http://www.uni-frankfurt.de/~wirth/>- Homepage of the International Research Group on Abductive Inference; at the Johann Wolfgang Goethe-Universität Frankfurt am Main, Germany.

Voelkel, J. R. (2001) *The Composition of Kepler's Astronomia nova,* USA-Princeton: Princeton University Press

Abduction as a strategy for concept formation in mathematics: Cardano postulating a negative.

Albrecht Heeffer
Center for Logic and Philosophy of Science
Ghent University
albrecht.heeffer@telenet.be

Peirce on abduction: a recapitulation

Peirce's definition of abduction has been understood and explained in two different meanings, coined by Magnani (2001) as *selective* and *creative* abduction. Selective abduction is the process of finding the right explanatory hypothesis from a given set of possible explanations. A common example of selection is medical diagnosis. In contrast, creative abduction *generates* the (right) explanatory hypothesis. As we are discussing concept formation, the abductive reasoning in our case study is of the creative type. We propose to view the formation of a new concept in mathematics as the hypothesis abduced to explain an anomaly. Peirce does not discuss concept formation as such but comes very close when he describes abduction as "the only logical operation which introduces new ideas" (1958, 5, 171). The existence of an anomaly is crucial in the cases we have studied, and is expressed by Peirce in similar wordings. He describes abduction as a form of inference motivated by the observation of a "surprising fact" (1958, 5, 188-9), reasoning initiated by "genuine doubt" and "genuine surprise" (1958, 5, 524), a motivation to break away from our habits "due to some novel experience" (ibid.). Abduction takes place when "we find ourselves confronted with some experience contrary to our expectations" (1958, 7, 36). We will show below

that Peirce's depiction of the mental state associated with abduction aptly describes the historical cases under discussion.

A last relevant aspect of Peirce's account of abduction is that the generated hypothesis must follow specific conditions to become the explanatory hypothesis: "Namely, the hypothesis cannot be admitted, even as a hypothesis, unless it be supposed that it would account for the facts or some of them" (1958, 5, 188-9). This necessitates some necessary connection between the novel hypothesis and the observed anomaly. The generated hypothesis delivers, at that time, the most adequate explanation for the "surprising fact".

We will now sketch the historical background in which the two original concepts emerged in algebra and why they can be considered as explanations for the observed anomalies.

The historical context

In 1545 Cardano published his *Ars Magna*, which is considered to be one of the landmarks in early algebra. The main cause for this recognition is the publication of the rules for solving certain types of cubic equations. While this is regarded as one of the major achievements of sixteenth-century algebra we believe Cardano's work deserves even greater credit for introducing three important conceptual innovations. All three have contributed to the emergence of symbolic algebra. The first is the introduction of operations between two linear equations in order to eliminate unknowns. Until then operations were only performed on a single equation. The second is the acceptance of negative solutions to linear and quadratic equations. The third is the first formulation of what would become imaginary numbers. These last two concepts are the subject of our analysis. They are introduced in one chapter called *de regula falsum ponendis* (the rule of postulating a negative). We can therefore assume that Cardano believes they share a common mechanism. This historical case allows us to expound the idea that this mechanism is one of abduction and plays a major role in the formation of these two new mathematical concepts.

Negative solutions in early algebra

Firstly, we have to underline that negative solutions to linear equations and negative roots of quadratic equations where completely unacceptable before Cardano. Those who have studied so-called negative solutions in early algebra (Sesiano 1985, Gericke 1996) failed to recognize that by introducing a debt in linear problems, Renaissance abacus masters explicitly deny negative solutions. Negative solutions were then a major oddity for which no sensible explanation could be given.

Algebraic problem solving in Renaissance Italy was imbedded in a rigid rhetorical structure, being both its major strength and its weakness. The rhetorical structure compensated for the lack of symbolism but prohibited, for example, multiple solutions to quadratic problems. Where two positive solutions for certain type of quadratic problems were accepted in Arabic algebra, their recognition gradually fades within the abacus tradition (Heeffer, 2006). The reason is that the analytical structure of the problem solving process poses one specific arithmetical value as the *cosa*, the unknown. When one poses the *cosa* for the smaller part of a division problem which reduces to a quadratic one, it makes no sense to arrive at a value for both parts.

Negative values in an algebraic solution to a problem first turn up in Europe in the work of Fibonacci. Apart from his *Liber Abbaci*, his best known work, Fibonacci wrote several other texts. In the *Flos*, several linear problems are discussed at length. A typical problem is that of men comparing their money:[1]

> There are four man each having a number of bezants. The first, with half of the bezants of the three others, has 33, the second, with a third part of the bezants of the others, has 35, again, the third, with one fourth of the bezants of the others, has 36, and the fourth, with one fifth of the bezants of the first, second and third, has 37. Asked is how much each of them has.

1 This and the following quotations are my translation from the Italian by Picutti 1983, 324-6, 365-6: "Vi sono dunque quattro uomini che hanno dei bisanti; il primo di loro con la metà dei bisanti dei rimanenti tre uomini, ne abbia 33; il secondo con la terza parte dei bisanti dei rimanenti tre ne abbia 35; ancora, il terzo con ¼ dei bisanti dei rimanenti ne abbia 36; il quarto poi, con 1/5 dei bisanti del primo e del secondo e del terzo uomo, ne abbia 37. Si domanda quanto abbia avuto ciascuno di essi".

This is a very old problem. Known commonly as the problem of men who want to buy a horse, its history is discussed by Tropfke (1980, 608) and Singmaster (2004, 7R.2). Early versions appear in Diophantus' *Arithmetica* and in Chinese and Hindu texts. A possible source of Fibonacci is Al-Karkhī (Woepcke 1853). However, the use of *bisanti* suggests an influence from a Byzantine source. Fibonacci was the first in the western world to apply algebra to the solution of the problem. In a modern reading, the problem can be represented by a set of linear equations using the unknown x, y, z and t for the shares of the four persons:

$$x+\frac{1}{2}(y+z+t)=33$$
$$y+\frac{1}{3}(x+z+t)=35$$
$$z+\frac{1}{4}(x+y+t)=36$$
$$t+\frac{1}{5}(x+y+z)=37$$

Such representation allows us to grasp the problem at once. With four equations and four unknowns, the problem is a determinate one which must lead to a unique solution. However, this symbolic representation is misleading for Fibonacci does not use these equations at all and it is inadequate to consider his solution method a resolution of equations. Furthermore, the solution method of the original text does not use four unknowns but a single one. At no point in the text are the shares of the four persons considered as algebraic unknowns. The problem is introduced with the following words:[2]

> I have fixed these numbers with great consideration, with the purpose that the solution of this problem falls in entire numbers; and I will show that with the given conditions it is insoluble.

2 Ibid. "Ho fissato questi numeri meditatamente, allo scopo che la soluzione di questo problema cada in numeri interi; e mostrerò che con le condizioni poste esso è insolubile".

Fibonacci starts by warning cardinal Ranerio Capocci, to whom the book is dedicated, that given the values, the problem has no proper solution, it is unsolvable. Looking for an integral solution he arrived at a negative value for one of the shares and this was difficult to accept at that time. Problems with similar solutions from Fibonacci's *Flos* and *Liber Abbaci* have been discussed both by Sesiano (1985) and Gericke (1996). Fibonacci must have been puzzled when he arrived at the value of a man's share being 33 – 36. He probably recalculated the whole procedure to double check for possible errors. In the end he found an interpretation for it. The man started with a debt to the other persons of 3 bezants. Sesiano and Gericke have no trouble interpreting the text as an instance of a solution with a negative number. We believe that a more cautious approach is appropriate here. The full acceptance of negative solutions to an equation was not established before the seventeenth century. Fibonacci explicitly avoids a negative solution in this text. He calls the problem unsolvable "unless one considers the first man having a debt". The interpretation as a debt actually disaffirms a negative solution. It must have bothered Fibonacci to such degree that he felt the need to apology before introducing the problem to the cardinal. He also looked for larger values to avoid assuming a debt. Probably by trial and error, Fibonacci tweaked the problem values to get rid of the "impossible solution":[3]

> And if we will say that the first man together [with the part of the others] has the required 181, the second 183, the third party 184, the fourth 185; all arranged as written, we will find that the first man has 1, the second 94, the third party 125, and he fourth 141.

$$x+\frac{1}{2}(y+z+t)=181$$

$$y+\frac{1}{3}(x+z+t)=183$$

$$z+\frac{1}{4}(x+y+t)=184$$

$$t+\frac{1}{5}(x+y+z)=185$$

3 Picutti 1983, 326: "E se diremo che il primo uomo abbia assieme al chiesto 181, il secondo 183, il terzo 184, il quarto 185, disposto tutto come sopra scritto, troveremo che il primo uomo ha 1, il secondo 94, il terzo 125, il quarto 141".

In fact, the revised problem as given by Fibonacci is the smallest integral solution for the given shares and the values in arithmetical progression n, $n + 2$, $n + 3$, $n + 4$, as shown by Sesiano (1985). Fibonacci must have spend a long time increasing the values until he arrived at four positive quantities. Such cases of negative values, and their interpretation as a debt, appear in isolated instances throughout the abacus texts of the next centuries. However, writing the history of negative numbers, Sesiano and Gericke look for the elements that support their story, and fall into the trap of biased observations. At the contrary, these cases support our claim that negative solutions were not accepted before the sixteenth century.

Before turning to Cardano's approach to negative values, we first sketch the context of using geometrical demonstrations for solving quadratic problems. This will allow us to understand Cardano's reasoning within the context of sixteenth-century algebra.

Geometrical demonstrations and quadratic problems

The algebraic solution of quadratic problems in medieval Europe was fully determined by the procedures of Arabic algebra. This knowledge was introduced in Europe through three Latin translations of al-Khwārizmī's *Algebra*, extant in sixteen manuscripts (Hughes 1982). These translations have been identified as from Robert of Chester (c. 1145), Gerard of Cremona (c. 1150) and Guglielmo de Lunis (c. 1215), although there is still discussion if the latter translation was to Latin or to Italian. What became available to the West was only the first part of al-Khwārizmī's treatise. Here, al-Kwārizmī gives solutions to algebraic problems by applying proven procedures in an algorithmic way. Early Arabic algebra recognized three types of quadratic problems, each having a given solution procedure. The validity of these solutions is demonstrated by geometrical diagrams. In contrast with Babylonian algebra the solution method is not geometrical in nature, only the demonstration is. In order to understand Cardano's approach we briefly discuss the geometrical demonstration of the case in which the *census* (or *subtancia* for Robert of Chester) and number equals a number of roots. The traditional example in symbolic representation is the equation

$$x^2 + 21 = 10x$$

This example appears in al-Khwārizmī's treatise (Robert of Chester, c. 1145, Karpinski 85; Hughes 1989, 39-41) as well as in the algebra of Abū Kamil (Levey 1966, 40; Sesiano 1993, 332) and many abacus texts of the fourteenth and fifteenth centuries. The geometrical demonstration is based on the representation in Figure 1.

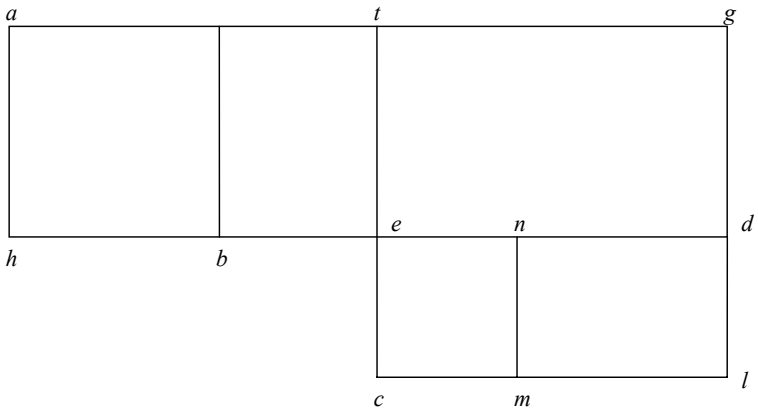


Figure 1: a geometrical proof for a quadratic problem from Arab algebra

The proof runs as follows:

- let x^2 be *ab* and $10x$ be *ad*
- therefore *ah* equals *x*, *hd* equals 10 and the area of *bg* is 21
- bisect *hd* at point *e*
- construct the square *tl* with side *tg*; its area is equal to 25
- subtract from *tl* the areas *tb* and *td* which results in the square *cn*
- this square has area 25 – 21 = 4, its side *ec* therefore is 2
- then *x* is *tc* minus *ec* or 5 – 2 equal to 3

So, for this type of quadratic problems, two positive solutions are recognized.

With this background, we can now turn to the chapter in which Cardano introduces some important new concepts in algebra.

Cardano posing a negative

Chapter 37 of the *Ars Magna* is named *de regula falsum ponendis*, or the "Rule of Postulating a Negative". Cardano introduces the subject by giving three reasons why it may be useful to postulate a negative.

- One can "assume a negative". The term "assuming" here refers to the rhetorical structure of algebraic problem solving. Every solution in abacus texts commences with the assumption that the rhetorical unknown equals some unknown quantity in the problem formulation. Assuming a negative thus means that one can use a negative *cosa*, or $-x$ instead of x, for the unknown quantity. This application is illustrated with three problems.
- One "seeks the square root of a negative".[4] After a failed geometrical demonstration Cardano solves two problems using imaginary numbers, for the first time in the history of mathematics.
- One can "seek what is not". With this cryptic formulation, Cardano refers to a negative which is neither of the first category nor of the second. However, his result stems from a wrong manipulation of imaginary numbers.[5]

We will discuss the first two applications of postulating a negative.

A negative solutions to a linear problem

Cardano illustrates the first use of postulating a negative with a problem which appears as a quadratic one, but reduces to a linear one (Cardano, *Opera Omnia* III, 286-7; Witmer 1968, 218):

---

4 Witmer (1968, 215) wrongly translates this as "one seeks a negative square root". Cardano 1663, III, 286, writes " $\mathfrak{R}$ m", meaning $\sqrt{-x}$.

5 In the last problem Cardano wrongly assumes that $\left(\frac{1}{4}\right)\left(-\sqrt{-\frac{1}{4}}\right)=\frac{1}{8}$. Operations with imaginary numbers can be tricky as shown by Euler (1770) in the example of the fallacy $-2=\left(\sqrt{-2}\right)\left(\sqrt{-2}\right)=\sqrt{(-2)(-2)}=\sqrt{4}=2$

> The dowry of Francis' wife is 100 aurei more than Francis' own property, and the square of the dowry is 400 more than the square of his property. Find the dowry and the property.
>
> We assume that Francis has – x; therefore the dowry of his wife is $100 - x$. Square the parts, making x^2 and
>
> $$10000 + x^2 - 200x$$
>
> The difference between these is 400 aurei. Therefore
>
> $$x^2 + 400 + 200x = 10000 + x^2$$
>
> Subtract the common terms and you will have 9600 equal to $200x$, wherefore x equals 48 and so much he has in the negative, i.e., is lacking, and the dowry will be the residue of 100, namely 52. Therefore Francis has – 48 aurei, without any capital or property, and the dowry of his wife is 52 aurei. By working this way, you can solve the most difficult and inextricable problems.

A straightforward transliteration of the problem into algebra would lead to

$$x^2 + 200x + 10000 = x^2 + 400$$
$$200x = -9600$$

in which – 48 is a valid solution to the equation. As this was still considered an anomaly in the early sixteenth century, Cardano assumes that the property is – x instead of x. By doing so, he arrives at a positive value 48 for the unknown quantity. As he posed that the quantity of the dowry was – x, the value thus becomes – 48. The Latin text reads "igitur Franciscus habuit 48 aureos debiti", which sounds alike the interpretation of a debt by Fibonacci to the 'negative solution' of a linear problem. However, now we do see the acceptance of a negative solution. In this and the next problem solutions, Cardano uses '1 *res* m.' for – x, and '4 m.' for the result – 4. The reason why negative values now become acceptable is related to the problem formulation. In order to make this more clear, let us go back to the traditional quadratic problems. As argued before, the rhetorical structure of abacus texts limits the solution of a quadratic problem to a single value.

The most common problem, known already in Babylonian algebra, is to find two numbers, given their sum (or difference) and their product. In modern symbolism this amounts to the general problem:

$$x + y = a$$
$$xy = b$$

In the abacus tradition, with some rare exceptions, all problems are solved using a single unknown. Depending on the initial assumption in the problem solution, the unknown represents one of the two unknown quantities. For example, the problem with the difference between two numbers being 7 and their product 60, could be represented by the symbolic equations

$$b = a + 7$$
$$ab = 60$$

If the *cosa*, x is used for the value b, then the quadratic equation is

$$x^2 - 7x - 60 = 0$$

with solution $x = 12$. On the other hand, if we use x for the value b, then the equation becomes

$$x^2 + 7x - 60 = 0$$

with the single solution $x = 5$. Arguing that a quadratic equation has two roots, therefore has little meaning in problem solving in the abacus tradition. The reformulation of the problem with a hypothetical unknown determines the solution than one will find.

Now, Cardano uses $-x$ as the rhetorical unknown. By "postulating a negative" he evidently arrives at a negative value. Embedded within the rhetorical structure, a negative solution now becomes perfectly acceptable. The acceptance of negative solutions is made possible by postulating negative unknowns.

Within two decades after the publication of Cardano's *Ars Magna*, isolated negative terms appear more frequently in algebra text books. The concept of a negative solution and consequently of negative algebraic terms are the

result of an abductive reasoning step. Negative solutions were considered an anomaly since their first appearance in European algebra with Fibonacci. However, because of an epistemological trust in the value of the algebraic procedures leading to these solutions, an explanation urges itself on the practitioners of early algebra. Cardano looks for an explanatory hypothesis to make a negative solution acceptable. He finds this abductive hypothesis in the analytical reasoning of algebra and the rhetorical structure of abacus texts. Both start by defining a hypothetical unknown for an unknown quantity of the problem. By hypothesizing a negative unknown, a negative value for the quantity becomes perfectly acceptable.

Imaginary numbers

With the second motivation for postulating a negative, Cardano again writes history. A similar hypothesis allows Cardano to apply the rules of signs to negative values inside the root and in this way be the first to partially formulate the operations on imaginary numbers. As pointed out by Leibniz (1712), imaginary numbers can best be defined by enumerating the symbolic calculations that are possible on those objects. This coincides with the historical emergence of the imaginary numbers. Incidentally, this is also the most common way in which secondary school students are first exposed to them.

The root of minus fifteen was an even greater abnormality than negative solutions. Cardano arrives at $\sqrt{-15}$ using standard problem solving methods. Devising a geometrical demonstration for the problem, the best qualification Cardano could come up with is "a missing surface". Confronted with this strange new concept, Cardano arrives at a strategy to make it, to a large extend, acceptable. He does so by finding the right abductive hypothesis which leads to the acceptance of the new concept, posing a negative value.

Adeo est subtile, ut sit inutile

Cardano was the first to perform a calculation with - what was later called - imaginary numbers. In the chapter we have just discussed Cardano (*Opera Omnia* III, 287; Witmer 1968, 219) writes:

The second species of negative assumption involves the square root of a negative. I will give an example: If it should be said, divide 10 into two parts the product of which is 30 or 40, it is clear that this case is impossible. Nevertheless, we will work thus: We divide 10 into two equal parts, making each 5. These we square, making 25. Subtract 40, if you will, from the 25 thus produced, as I showed you in the chapter on operations in the sixth book, leaving a remainder of – 15, the square root of which added to or subtracted from 5 gives parts the product of which is 40. These will be

$5+\sqrt{-15}$ and $5-\sqrt{-15}$.

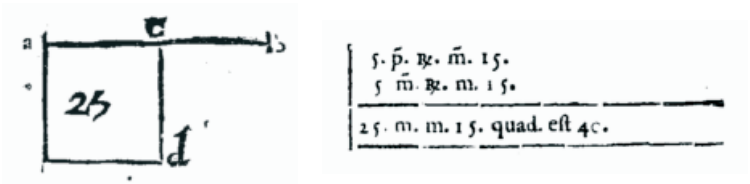


Figure 2: First occurrence of imaginary numbers from Cardano's Ars Magna (1545)

Using a geometrical demonstration he tries to get a grasp on the new concept (Figure 2). He proceeds as with the standard demonstration of the rule for solving quadratic problems. Let *AB* be the 10 to be divided. Divide the line at *C* into two equal parts. Square *AC* to *AD*. Since 40 is four times 10 this corresponds with the rectangle 4*AB*. Now $\sqrt{-15}$ corresponds with the subtraction of *AD* by the larger 4*AB*. Thus, Cardano finds that this strange new object is a negative or a missing surface. This makes no sense to him and therefore he writes that the problem is impossible.

Cardano still struggled with the interpretation of $\sqrt{-15}$, but was "putting aside the mental tortures involved" and performed the operation $(5+\sqrt{-15})(5-\sqrt{-15})$ correctly to arrive at 25 + 15 or 40. Multiplying the two binomials produces four terms. The first is evidently 25. The second and third are cancelled out by their signs.

$$5\left(\sqrt{-15}\right)+5\left(-\sqrt{-15}\right)$$

The innovation lies in the fourth.

$$\left(\sqrt{-15}\right)\left(-\sqrt{-15}\right)=-(-15)$$

Remark that the only way to accept these operations for the Renaissance mind is to approach them as purely symbolical. By replacing the – 15 with x as in "posing a negative" and you get a common operation:

$$\left(\sqrt{x}\right)\left(-\sqrt{x}\right)=-x$$

Cardano is able to accept the operations as valid *on the basis of their symbolic structure*, although he does not know what it means. Cardano closes the discussion with the words that this is a kind of arithmetic subtlety, "as refined as it is useless" ("adeo est subtile, ut sit inutile"). However, he is convinced that the operations are correct.

Concept formation by abduction

In both cases Cardano was confronted with an anomaly which appeared for several centuries of algebraic practice. Rather than avoiding such problems or keeping them in the dark, Cardano succeeded in finding a satisfactory solution, which he called "postulating a negative". He constructed an explanatory hypothesis which made these "surprising facts" acceptable. By doing so he created new concepts, such as imaginary numbers, which eventually lead to a whole new branch within mathematics. Both cases provide historical evidence of the emergence of new concepts in mathematics by abductive reasoning.

The question we can now ask is if there would have been other ways to deal with these anomalies at the time of Cardano. Most probably there were. In some way, the interpretation of Fibonacci of a negative solution as a debt is a form of abductive reasoning which also resolves the issue. It is, however, not an interesting one from a mathematical point of view. Also for imaginary numbers there may have been alternatives. In a recent paper, Jean Paul van Bendegem argued that a geometrical interpretation operating on missing surfaces could have provided an alternative to the symbolic model as it has emerged.[6] van Bendegem applies his geometrical interpretation to the so-called *casus irreducibilis*

6 van Bendegem, unpublished, "What could an 'alternative' mathematics be like, if anything?", presented at the ninth VlaPoLo conference, 6-8 May 2005, Ghent.

$$x^3 = px + q \quad \text{with} \quad \left(\frac{1}{3}p\right)^3 > \left(\frac{1}{2}q\right)^2, \quad \text{e . g .} \quad x^3 = 15x + 4 \ ,$$

the case studied by Bombelli (1572). Though he was well aware of the problem, Cardano quietly dismissed the irreducible cases in the first chapters of the *Ars Magna*. In that way, imaginary numbers are irrelevant for Cardano's approach to the cubic. But the geometrical interpretation proposed by van Bendegem is perfectly applicable for the quadratic problem discussed above.

However, we would like to argue that Cardano's approach was indeed the best one available at that time. The period of 1515 tot 1545 was one of the increasing use of symbols in mathematics. In fact, the plus and minus sign were introduced within the context of algebra in early sixteenth-century German manuscripts.[7] Once algebraic terms are affected by symbols, they turn into tools much more powerful. The use of symbols allows to abstract from the arithmetical content they represent. New operations become possible because they can be applied symbolically, without accounting for the values they stand for. Even if the idea of a negative value or the root of a negative value is still ridiculous, it now becomes possible to reason with negative terms on a symbolic level. As shown above, symbolic abstraction plays an important role in both cases of abductive reasoning by Cardano.

Acknowledgements

This paper is derived from chapter 9 of my PhD dissertation, *From Precepts to Equations: The Conceptual Development of Symbolic Algebra in the Sixteenth Century*, supervised by Diderik Batens, Ghent University Belgium. Funding for this research was provided by the research project G.0193.04 from the Flemish fund for scientific research (FWO Vlaanderen). I would like to thank Joke Meheus for the discussion we had about these case studies.

7 The first appearance of the + sign in the Dresden codex C80, f. 350^v, written around 1486. Exemplary is the anonymous Vienna codex 5277, written between 1500 and 1518. Here the rules of signs are introduced with symbols as operations on polynomials.

References

Bombelli, Rafael (1572) *L'algebra: parte maggiore dell'arimetica : divisa in tre libri*, Giovanni Rossi, Bologna.

Cardano, Girolamo (1539) *Practica Arithmetice*, Bernardini Calusci, Milan.

Cardano, Girolamo (1545) *Ars Magna*, Johann Petreius, Nürnberg, (English translation by Witmer, R. T., 1968, *Ars Magna or the Rules of Algebra*, M.I.T. Press, Cambridge, Mass., Reprinted by Dover Publications, New York, 1993).

Cardano, Girolamo (1663) *Opera omnia* (10 vols.), Jean Antoine Huguetan and Marc Antione Ravaud, Lyon.

Euler, Leonard (1770) *Vollständige Anleitung zur Algebra* (2 vols.), Gedruckt bey der Kays. Acad. der Wissenschaften, St. Petersburg.

Gericke, Helmuth (1996) "Zur Geschichte der negativen Zahlen", in J. Dauben, M. Folkerts, E. Knobloch and H. Wuβing (eds.) (1996) *History of Mathematics. States of the Art*, Academic Press, New York, pp. 279-306.

Heeffer, Albrecht (2006) "A Conceptual Analysis of Early Arabic Algebra" to appear in Tony Street, Shahid Rahman and Hassan Tahiri (eds.) *Arab Logic And Epistemology*, Kluwer Academic Publishers.

Karpinski, Louis Charles (1915) *Robert of Chester's Latin Translation of the Algebra of Al-Khwarizmi*, Macmillan Company, New York (Reprinted by University of Michigan Press, Ann Arbor, 1933).

Leibniz, G.W. (1712) "Observatio, quod rationes sive proportiones non habeant locum circa quantitates nihilo minores, & de vero sensu methodi infinitesimalis", *Acta eruditorum*, pp. 167-9.

Levey, Martin (1966) *The algebra of Abu Kamil. Kitab, fi al-jabr wa'l- muqabala in a commentary by Mordecai Finzi*, The University of Wisconsin Press, Madison.

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

Peirce, Charles Sanders (1958-1966) *The Collected papers of Charles Sanders Peirce*, (8 vols.) Charles Hartshorne and Paul Weiss, (eds.), A.W. Burks (ed, vols. 7-8), Belknap Press of Harvard University Press, Cambridge.

Picutti, Ettore (1983) "Il 'Flos' di Leonardo Pisano dal codice E.75 P. Sup. della Biblioteca Ambrosiana di Milano", *Physis* 25, pp. 293-387.

Sesiano, Jacques (1985) "The appearance of negative solutions in Mediaeval mathematics", *Archive for History of Exact Sciences*, 32, pp. 105-150.

Singmaster, David (2004) *Sources in Recreational Mathematics, An Annotated Bibliography*, Eighth Preliminary Edition (unpublished, electronic copy from the author).

Tropfke, Johannes (1980) *Geschichte der Elementar-Mathematik in systematischer Darstellung mit besonderer Berücksichtigung der Fachwörter*, Bd. I *Arithmetik und Algebra*, revised by Kurt Vogel, Karin Reich and Helmuth Gericke, de Gruyter, Berlin.

Is the major conceptual moment of orthodox quantum mechanics, a moment of abduction?

João Cordovil
Centro de Filosofia das Ciências da Universidade de Lisboa
jlcordovil2@hotmail.com

In the beginning of the 20's of the last century the new born quantum physics faced a crisis. Despite its early great achievements, namely Einstein's explanation of the specific heat of solids and above all the formulation of Bohr's atomic model, quantum theory has accumulated anomalies and became more and more complicated.

This crisis was deeply conceptual, and necessarily connected with two great mysteries. On one hand, the atomic model of Bohr-Sommerfeld, which explained so well the hydrogen spectrum, was based on the assumption that only certain preferred motions, *quantized* motions, were possible or at least stable for the electron inside of the atom. This assumption was the central point of the model. However, it was unjustified and seemed rather strange and even ad-hoc. After all, how to explain that among an infinity of motions which the electron may have in the atom according to classical concepts, only certain ones were possible? By the other hand, the nature of light was still involved in obscurity. In the XIXth century physicists thought that light was a wave. A lot of experimental data seemed to show clearly that light had to be or to act like a wave. But in the beginning of the 20th century it also became clear that some phenomena were very difficult to explain without the assumption that light was constituted by discrete particles of energy. In other words, constituted by quanta. So, it seemed necessary to assume that light had the capacity of acting sometimes like a wave and sometimes like a

corpuscle. But how to understand this? How could this strange light's corpuscular behaviour be explained?

These two deep mysteries were installed in the main core of quantum physics. And they seemed completely unrelated. One dealing with matter, the other with light. One dealing with mechanics, the other with optics. And the only thing they apparently had in common was that both presented quanta features.

Without any doubt, a new quantum physics was needed. A quantum physics that could solve these two enigmas: to relate the wave behavior and the corpuscular behavior of light and that, at that at the same time, to justify the quantified orbits of the atom.

The appearance of this new quantum physics - the so called quantum mechanics - and the solution to these two enigmas was achieved by De Broglie's in 1924. How did he solve those mysteries? What was his reasoning?

As De Broglie stated in his Nobel Lecture, in 1929, his reasoning was the following:

> "When I started to ponder these difficulties two things struck me in the main. Firstly the light-quantum theory cannot be regarded as satisfactory since it defines the energy of a light corpuscle by the relation $E = h\nu$ which contains a frequency. Now a purely corpuscular theory does not contain any element permitting the definition of a frequency. This reason alone renders it necessary in the case of light to introduce simultaneously the corpuscle concept and the concept of periodicity.
>
> On the other hand the determination of the stable motions of the electrons in the atom involves whole numbers, and so far the only phenomena in which whole numbers were involved in physics were those of interference and of eigenvibrations. That suggested the idea to me that electrons themselves could not be represented as simple corpuscles either, but that a periodicity had also to be assigned to them too.
>
> I thus arrived at the following overall concept which guided my studies: for both matter and radiations, light in particular, it is necessary to introduce the corpuscle concept and the wave concept at the same time. In other words

> the existence of corpuscles accompanied by waves has to be assumed in all cases.
>
> […] Thus to describe the properties of matter as well as those of light, waves and corpuscles have to be referred to at one and the same time. The electron can no longer be conceived as a single, small granule of electricity; it must be associated with a wave"[1]

Based on this hypothesis - that both light and matter had to be described as something corpuscular associated with a wave - De Broglie was able to justify the quantification of the atom. He could also explain the strange dualist behaviour of light. And - this is the more amazing – he could predict the existence of matter waves, which were unknown at the time. In 1927, with the experimental observation of this matter waves, in 1927, De Broglie's hypothesis was confirmed and considered valid.

Now, the question we would like to point is: was De Broglie's discovery of matter waves made by some kind of an abduction reasoning?

In the majority of books of quantum physics, it is often suggested that De Broglie achievement has been made by induction. He would have simply generalized the Einstein's light-dualistic hypothesis to all quantum objects, namely, to electrons. Something like this: if it is the case for photons, then it will be the case for all quantum objects. Thus, according to this traditional view, De Broglie inductively concluded that the dualism of light was not a property of light, but a general law of nature. And, for that so, he deduced that every electron must have an wave associated.

That is to say: from this point of view, De Broglie "simply" extended the light quanta hypothesis to matter. But things are not so "simple". In fact, this De Broglie-inductive thesis seems to me to have two great problems.

First, until 1924, no one - not even Einstein - really believed that light had a dualistic wave-corpuscular nature. For the majority of physics, the quantum hypothesis was considered a practical tool to understand some strange phenomena. But nothing more than a useful tool.

1 See http://nobelprize.org/nobel_prizes/physics/laureates/1929/broglie-lecture.pdf

Thus, it cannot be argued that De Broglie simply extended the quantum-light hypothesis of light to electrons. Firstly he had to assume that the quantum-light hypothesis was correct. It was the "real" thing. And doing so, he was strongly compromising himself with an hypothesis that was not new, that could explain the experimental data, but no one truly believed it.

On the other hand, it seems to me that the main De Broglie's hypothesis was not, as the traditional view claims, that matter, like light, must have a dualistic nature. The main De Broglie's hypothesis was that there was a very strong analogy between light and matter. Or, in other words, that light and matter were not distinct or merely connected. But there were two different modes of the same Nature: the quantum nature. That is to say: at the quantum level, matter is like light, and light is like matter. And this was an astonishing hypothesis that couldn't be made by De Broglie using either an inductive reasoning, or a deductive reasoning.

Of course, one can argue that he simply used an analogy between light and matter. And this is true. But the problem is that this can never be considered as a simply analogy. Matter and light were, in physics, opposite. They were like water and fire. Or, like solid and fluid. So, it is true that De Broglie used an analogy. A so strange analogy that we have to ask: what kind of reasoning has he used in order to accomplish such analogy?

It seems to me that we can suppose that this major conceptual moment of quantum mechanics, its very born moment, was probably achieved by two great steps of abductive reasoning. First, the formulation of the generalized wave-corpuscular dualistic hypothesis. And, secondly, the synthesis between light and matter.

This is, of course, an open question. Nevertheless, an important one. Especially because the same question could be (can be) asked about the other main conceptual moments of quantum mechanics, namely Bohr's complementary principal and the formalisation of Schrödinger equation. And, for that so, it can be important to analyse carefully the role of abduction in the formulation of quantum mechanics.

Abductive reasoning, theoretical preduction , and the physical way of dealing fallibly with nature[1]

Andrés Rivadulla
Departamento de Lógica y Filosofía de la Ciencia
Universidad Complutense
arivadulla@filos.ucm.es

1. Introduction

Many philosophers of science think that science does harbour some privileged method of discovery. Were this right, then it would be possible for scientists to discover both truths about facts and truths about theories. Were science successful with both forms of dealing with Nature: the discovery of facts and the postulation of true theories about the world, it would be justified to claim that description, explanation, understanding and prediction are the aims of science.

The question is whether it really exists a privileged form of discovery in scientific methodology. From its very beginning nearly 2400 years ago, Western science has been concerned with two main forms of dealing scientifically with Nature:

1. According to Plato's astronomer method -as defined by Duhem (1908)- the aim of science is to propose hypotheses intended to save the appearances. The task of hypothesis building is provided by abduction.

2. According to Aristotle, the aim of demonstrative science is to discover the principles which would serve as explanations for the observed phenomena. The task of discovery is provided by *induction.*

1 This paper is part of a research project on *Theoretical Models in Physics* supported by the Spanish Ministry of Education and Science.

In the following sections I dwell upon induction and abduction, and conclude that they are heuristics providing hypotheses intended to deal fallibly with Nature. In the last section I introduce preduction, an implementation of the hypothetical-deductive method very commonly applied in theoretical physics. I will also argue that preduction is a reasoning strategy that merely allows us to deal fallibly with Nature. Fallibly dealing with Nature is all what science can do.

2. Induction

Let us assume *induction* to be a kind of ampliative (and truth conservative) inference. Then *inductive reasoning* would be a method of scientific discovery. In order to see if this holds true, I am going on to scrutinize Aristotle's view on induction.

In the Prior Analytics, Book II, $68^{b}15$-29: Aristotle proceeds in the following way:

Let *A* stands for long-liver,
B for bileless,
C for the particular long-lived animals, e.g. man, horse, mule.

And let "Induction (…) consists in proving through *C* that *A* belongs to *B*. For this is the manner in which we make inductions", then Aristotle's *inductive syllogism* looks like this:

All *C* is *A*	(*A* belongs to the whole of *C*)
All *C* is *B*	(*B* also belongs to every *C*)
Hence: All *B* is *A*	(*A* belongs to *B*)

Now, were this way of reasoning formally correct, then *after a finite enumeration of individuals sharing properties A and B*, it would be logically legitimate to conclude that *A* belongs to *B*. Thus Aristotle seems to assume that induction consists in proving *through* a finite class *C* of individuals, that *A* belongs to *B*.

Two problems arise immediately: The first one is that Aristotle's inductive syllogism is not a syllogism of the first figure (i. e. it is not like *Barbara*). The second is that Aristotle's inductive syllogism is not a valid syllogism of the third figure, i. e. it is not of the kind of *Darapti, Felapton, Disamis, Datisi, Bocardo, Ferison*. In conclusion: Aristotle's inductive syllogism is not any syllogism at all.

Aristotle was perfectly aware of this logical problem. Therefore he offered the following solution to it: "If then *C* is convertible with *B*, and the middle term is not wider in extension, it is necessary that *A* should belong to *B*." According to this solution, Aristotle's inductive syllogism would look like a syllogism in *modus Barbara*, and would be formally correct:

All *C* is *A*	(man, horse and mule are long-lived animals)
All *B* is *C*	(All bileless animals are man, horse and mule, i.e. the only bileless animals are man, horse and mule)
Hence: All *B* is *A*	(All bileless animals are long-liver)

Unfortunately, Aristotle's purported solution brought a new problem into light, since in order to affirm that *All B is C*, it is necessary to make a complete enumeration of all (species of) animals. Thus *Aristotle's cost for the intended conversion is complete enumeration*. Surprisingly Aristotle justifies his step by claiming (*Prior Analytics, Book II*, $68^{b}15$-29) that "we must apprehend *C* as made up of all the particulars. For induction proceeds through an enumeration of all the cases."

Were Aristotle prepared to accept that induction proceeds by complete enumeration of all cases, the cost would be to transform induction into deduction, thus ceasing to be a peculiar form of scientific discovery. The alternative could be to maintain induction as a logically illegitimate way of reasoning (thus there would not be any logically legitimate *inductive syllogism* at all). In this case inductive conclusions would not necessarily be true, and induction would become only *a fallible way of reasoning about Nature*.

Can abduction replace induction as a way of scientific discovery?

3. Abduction

Abduction is generally conceived of as a kind of reasoning by means of which surprising phenomena suggest hypotheses that may be used as premises for further predictions. The very well-known Peircian scheme of abduction is:

Some phenomenon *F* has been observed
Were *H* a true explanation of *F*, then *F* would be a matter of fact
Hence: There is reason to believe that *H* is true.

There are lots of examples of abduction in the history of science, of which a few ones are:

- The theoretical models accounting for planetary motions in ancient astronomy until Copernicus.
- The celebrated abduction by Kepler of Mars's elliptic orbit.
- Charles Darwin's hypothesis of natural selection in *On the Origin of Species by Means of Natural Selection*, 1859, abductively inferred on the basis of abundant available biological and fossil evidence.
- Ernest Rutherford's postulation of the *atomic planetary model* in 1911 on the basis of the scattering of alpha-particles by a thin gold plate in an experiment conducted by Geiger and Marsden in 1910.
- Alfred Lothar Wegener's (1880-1930) postulation of the hypothesis of *continental drift* in his *Die Entstehung der Kontinente und Ozeane* (*The Origin of the Continents and Oceans*), 1915. Evidences supporting Wegener's abduction of the *continental drift hypothesis* were:

1) Measures of the continuous separation of Europe and America;
2) Geographic-morphological coincidences between South-America and Africa, and between Labrador, Greenland and Scandinavia.
3) Geological affinities of the cordilleras in Norway, Scotland and Canada; of the plateaus of Africa and Brazil; and between Patagonia and the Cape region.
4) Paleontological discoveries of fossils of pre-historical identical organisms in distant regions of South-America, India and Australia

5) Paleoclimatic signs: In Pangea the Poles didn't coincide with the actual ones. Some tropical and subtropical regions today were covered with ice 300 millions years ago. And vice-versa, tropical regions then are situated today in very cold zones, as fossil petrol and coal deposits show.

No wonder that abduction has attracted the interest of so many philosophers of science, who have seen in it a form of *scientific discovery,* an *inference to the best explanation*, or merely a tool for *scientific creativity*. According to Lorenzo Magnani (2001, 17-18)

> Theoretical 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.

And for John R. Josephson (1994, p. 5):

> *Abduction*, or *inference to the best explanation*, is a form of inference that goes from data describing something to a hypothesis that best explains or accounts for the data. Thus abduction is a kind of theory-forming or interpretative inference.

Abductions share with inductive generalizations (Josephson, p. 13), that they

> are *ampliative inferences*; that is, at the end of an abductive process, having accepted a best explanation, we may have more information than we had before. The abduction transcends the information of its premises and generates new information that was not previously encoded there at all.

But abductions disagree with deductions that

> Deductions are *truth preserving*, whereas successful abductions may be said to be *truth producing*.

My question is: *Is abduction really an inference to the best explanation?* The first method of abduction in the history of science goes back to Plato. Plato's *astronomer method* is implicit in the following question (Cf. Duhem 1908, 5):

> What circular motions, uniform and perfectly regular, are to be admitted as hypotheses so that it might be possible to save the appearances presented by the planets?

According to Duhem's (1908, 5-6) interpretation of Plato's methodological proposal,

> 1. Astronomy is the science that so combines circular and uniform motions as to yield a resultant motion like that of stars. When its geometric constructions have assigned each planet a path which conforms to its visible path, astronomy has attained its goal, because *its hypotheses have then saved the appearances.*
>
> 2. The astronomer must declare himself fully satisfied when the hypotheses he has combined succeed in saving the appearances[2].

Duhem's interpretation thus sheds some doubts on the view of Platonic abduction as inference to the best explanation. (And Plato's *abductive reasoning* taken by and large provided the way for the construction of geometrical models in ancient astronomy until Copernicus.) Moreover, according to Peirce (C.P. 2.776 and 2.777)

> the hypothesis which [presumption, abduction, A.R.] problematically concludes is frequently utterly wrong itself, and even the method need not ever lead to the truth.

Since Peircian abduction shares with induction that any abduced conclusion can be wrong, should abduction be conceived of as a method of discovery, or merely as a kind of heuristics providing hypotheses intended to deal fallibly with Nature?

2 This sentence reminds Paul Dirac's (1963, 48) view that "If the physicist knows how to calculate results and compare them with experiment, he is quite happy if the results agree with his experiments, and this is all he needs."

4. Modeling abduction from the viewpoint of Bayesian probabilistic inference

According to Josephson's conceptual analysis (*op. cit.*, p. 5), abduction is a distinctive kind of inference that follows this pattern:

> D is a collection of data (facts, observations, givens).
> H explains D (would, if true, explain D).
> No other hypothesis can explain D as well as H does.
> Therefore, H is probably true.

Now, if abduction is inference to the best, i. e., the most probably true explanation, then there are to ways, both of them put forward by Josephson (p. 14), by which abduction could be strengthen:

1. an abductive process might seek further information beyond that given in the data initially to be explained.
2. there may be a need to distinguish between explanatory alternatives; for help in forming hypotheses; or for help in evaluating them.

Josephson's case 1. above amounts to determining whether or not some growing evidence on behalf of a hypothesis -further information beyond that given in the data initially to be explained- contributes to increase its probability to be true.

Let e_i, e_j, e_k, e_l be different evidences favorable to a given hypothesis h. And let assume that each one of these evidences satisfy the following conditions: 1) $h \rightarrow e_i$ and 2) $0<p(h)<p(e_i)<1$. If the evidence supporting h growths, Bayes' Theorem looks like:

$$p(h \mid e_j \wedge e_i) = \frac{p(h)\, p(e_j \wedge e_i \mid h)}{p(e_j \wedge e_i)}.$$

Well, it happens that

$$p(h \mid e_j \wedge e_i) > p(h \mid e_i).$$

And in general:

$$... > p(h \mid e_j \wedge e_i \wedge e_k) > p(h \mid e_j \wedge e_i) > p(h \mid e_i) > p(h)$$

Now, two problems arise here:

1. When either novel evidence or even predicted evidence continues to support the hypothesis, then, according to Bayes' theorem, the probability for the hypothesis to be true would increase continuously. Thus after a considerable amount of favorable evidence the probability of the hypothesis to be true would approach 1. Nevertheless, if *surprisingly new evidence or a new prediction of the hypothesis fails, then a highly probably true hypothesis suddenly would become false.*
2. Newtonian mechanics was a highly confirmed hypothesis. But let's write its history differently: Let us assume that Newton's *Query I* of his *Optics*, 1704: "Do not bodies act upon light at distance, and by their action bend its rays; and is not this action (*coeteris paribus*) strongest at the least distance?" had been answered earlier: the bending angle amounts to 0".87 seconds of arc[3]. And let assume that Newton himself had given the conditions for the confirmation of this prediction, and that astrophysical and technical conditions would have been given for the empirical proof of the hypothesis. *Then Newton himself would have been forced to recognize that his gravitational hypothesis was simply wrong.*

Josephson's case 2. above amounts to determining whether an available evidence can contribute to discriminate among many competing hypotheses: "there may be a need to distinguish between explanatory alternatives".

It seems here that abduction is a form of Baconian eliminative induction, and Bayes' Theorem can be used as well as a model for it. A Bayesian model for Case 2. might be the following: Let the empirical evidence *e* be the angular deviation $\Delta\vartheta = 1''.75$ of light by the Sun. And let the alternative explanatory hypotheses be relativity theory (*TR*), Newtonian mechanics (*MN*), and a third hypothesis h_3, which is assumed to be theoretically incompatible, but empirically equivalent, with *TR*.

Bayes' Theorem takes the form

$$p(h_i \mid e) = \frac{p(h_i)\; p(e \mid h_i)}{\sum_{i=1}^{3} p(h_i)\; p(e \mid h_i)},$$

3 Cf. A. Rivadulla (2004a).

and its applicability is restricted by conditions: $p(h_1)+...+p(h_n)=1$, and $p(h_1|e)+...+p(h_n|e)=1$.

We know that according to *MN* the value of the light deflection should amount to 0".87, whereas *TR* predicts a deviation of 1".75, and the same does h_3, which is assumed to be empirically equivalent to *TR*. Thus, given that the available evidence is $\Delta\vartheta = 1''.\mathfrak{Z}$, then $p(\vartheta = 1''.\mathfrak{Z} \mid \mathit{MN}) = 0$ and $p(e|h_3) = p(\vartheta = 1''.\mathfrak{Z} \mid \mathit{R}) = 1$.

Let's now distribute the initial probabilities of these hypotheses in the following way (These values are arbitrary, but irrelevant): $p(MN)=0.5$, $p(TR)=0.35$, $p(h_3)=0.15$. Applying Bayes' Theorem it results that $p(MN|e) = 0$, $p(TR|e)=0.7$, and $p(h_3|e)=0.3$.

At the first look Bayes' Theorem seems to accomplish its task: The probability of Newtonian Mechanics to be true becomes 0, since it has been refuted by experience, whereas the probability of both relativity theory and h_3 has grown, since they have both been confirmed by experience.

But this last claim is misleading, since as following equivalence shows

$$\frac{0.35}{0.15} = \frac{0.7}{0.3}$$

the reason between a priori probabilities holds also a posteriori[4].

Thus Bayes' Theorem is of no help in modeling abduction as an inference to the best, i. e. to the most probably true hypothesis. Abduction remains a kind of heuristics providing ways of reasoning that merely allow us to infer theoretical constructs intended to deal fallibly with Nature.

5. Theoretical prediction

Induction and abduction clearly fail as legitimate logic forms of scientific discovery. Thus there seems to be *stricto sensu* no *ars inveniendi* in scientific methodology capable to lead to true (or even probably true) theories, theoretical models or theoretical explanations.

Nonetheless induction and abduction are not the only ways of reasoning capable to produce hypotheses usable for further fallible inferences.

4 This example has been taken from Rivadulla (2004*b*, Chapter 1)

Astonishingly theoretical *prediction* has escaped notice in the philosophy of science[5], although it is a methodological strategy theoretical physicists commonly apply in order to postulate theoretical models, hypotheses, laws and principles.

I propose here to call *prediction* a form of reasoning that starts *ab initio* from first principles, *methodologically* postulated as premises of an inferential procedure. Predictive reasoning differs from abduction in that the preduced products are not suggested by data, but constructed on the basis of the available theoretical background. (This is the reason why I call it *theoretical prediction*.) Thus they depend more on the theoretical framework than on the empirical data. Prediction implements the hypothetical-deductive method, but it should not be confused with the axiomatic-deductive method.

The reason for the differentiation of this particular implementation of the hypothetical-deductive method, which I call prediction, is that the premises used for further inferences in a predictive argument very usually proceed from different theories. This means that physicists resort to the available theoretical background *as a whole* in order to achieve their inferential aims. Substitution *salva veritate* and dimensional analysis are important means to reach their aims. Only one strategy is forbidden: to use theories in parts where they contradict each other.

Prediction provides indeed the way of reasoning by which most theoretical models are postulated in science, mainly when we are concerned with domains orphan of theories. This is the case of astrophysics. In astrophysics scientists make use of the whole of physics, in order to accommodate the scarce data provided by the light of the celestial objects astrophysicists are concerned with. But prediction does not only provides us with theoretical models; factual hypotheses and theoretical laws are also theoretical products achieved by the application of predictive reasoning. Factual hypotheses like the temperature of the solar surface, Jeans mass critical value for the formation of protostars, or the critical density of the Universe; theoretical laws like Rayleigh-Jeans law, Planck's own radiation law, or Schrödinger's equation;

5 As I have been told, the word *prediction* has been used previously in artificial intelligence by Jum Arima and by Allen Courtney and Norman Foo.

and theoretical models, like Helmholtz-Kelvin gravitational collapse model for the energy of stars, Bohr's atomic planetary model, or the stars interior models in astrophysics, are some examples of theoretical products achieved by preduction in physics.

Nevertheless a last important point needs to be made clear. Since preduced theoretical constructs depend on the available theoretical background, and this cannot be known to be true, restrictions do frequently occur in the domain of their intended applications, and refutations sometimes happen as well. We learn indeed from the methodology and the history of theoretical models in physics, for instance, that no matter how a given theoretical model has been achieved: by induction, by abduction or even by *preduction*, we cannot a priori know how many phenomena will belong to its application domain, nor how many of them will have to be removed from its domain of intended applications[6]. *Thus preduction only supplies a fallible strategy in order to deal predictively with Nature.*

Induction, abduction and preduction are merely fallible ways of dealing scientifically with Nature. This is all what theoretical science can do.

References

Aristotle, *The Complete Works*. The revised Oxford translation. Edited by Jonathan Barnes, Princeton University Press, 1984

Dirac, P. (1963): "The Evolution of the Physicist's Picture of Nature". *Scientific American*, Vol. 208, Number 5, 45-53

Duhem, P. (1908), *To Save the Phenomena. An Essay on the Ideal of Physical Theory from Plato to Galileo*. Hermann et Fils, Paris 1908. The University of Chicago Press, Chicago 1969

Josephson, J. R. (1994): "Conceptual analysis of abduction". In John R. Josephson and Susan G. Josephson (eds.), *Abductive inference. Computation, Philosophy, Technology*, Cambridge University Press, New York

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

6 Cfr. Rivadulla (2006, Sections 3 and 4)

Newton, I. (1704), *Optics*. Encyclopaedia Britannica, 4th printing 1993, Chicago

Peirce, C. S., *Collected Papers*, Harvard University Press, Cambridge, MA, 1931-1958

Rivadulla, A. (2004*a*): "The Newtonian Limit of Relativity Theory and the Rationality of Theory Change". *Synthese* 41, 417-429.

Rivadulla, A. (2004*b*), *Éxito, Razón y Cambio en Física. Un enfoque instrumental en teoría de la ciencia.* Editorial Trotta, Madrid.

Rivadulla, A. (2006): "The Role of Theoretical Models in the Methodology of Physics". In L. Magnani (ed.), *Model-Based Reasoning in Science and Engineering* (forthcoming)

4

ABDUCTIVE STRATEGIES, APPLICATIONS & PERSPECTIVES

Seeing problems, seeing solutions. Abduction and diagrammatic reasoning in a theory of scientific discovery

Michael H.G. Hoffmann
Georgia Institute of Technology
m.hoffmann@gatech.edu

Introduction

In a recent paper, Hanne Andersen describes how physicist Ida Noddack formulated in 1934 for the first time the hypothesis that an experiment described by Enrico Fermi the same year could be explained as a case of nuclear fission (Andersen 2006). Surprisingly, the scientific community at this time did not even discuss this farsighted, and for modern physics absolutely fundamental explanation. It needed four years until Hahn and Straßmann raised once more the hypothesis that in this experiment a nucleus had split into fractions which was then immediately welcomed as the discovery of nuclear fission.

There are two questions that are interesting here with regard to the general problem of the possibility, and the process, of scientific discoveries:

1. Why was Noddack's suggestion ignored?
2. How was it possible, that Noddack *saw* something in the data that others did not see?

Andersen shows that the scientific community was divided into two factions in 1934. The vast majority of scientists assumed that artificially induced "disintegration processes *had* to be one nucleus transmuting into another nucleus of almost the same size by releasing a small particle. On this model, there was no way a nucleus could divide into a few large fractions" (2006, p. 16). Noddack, on the contrary, was less constrained by physical consid-

erations regarding a generally accepted taxonomy of possible disintegration processes. She "was an analytical chemist who had worked for years searching for the still missing elements in the periodic table" (p. 23). Therefore, she focused first of all on the chemical properties of what has been produced by Fermi's experiment. While Fermi assumed that what he produced by bombarding uranium with neutrons was the transuranic element of atomic number 93, the chemical properties the produced element displayed were different from what was expected by chemists for element 93. Andersen shows that, for Noddack, "chemical identifications clearly had much more weight in identifying the transmuted nuclei than physical expectations of possible decay series" (p. 24). This means that she felt a severe anomaly where "no anomaly was seen" by the majority of scientists; "and without a serious anomaly, there was no reason to accept a radical change of a highly successful taxonomy" (ibid.). Such a change was only possible for the physicists when by the end of 1938 also in another, well-investigated part of the original taxonomy a serious anomaly was discovered. Now there was "a gap in the taxonomy ..., ready to be filled" by the "new" phenomenon of nuclear fission (ibid.).

What I want to highlight with this example is that *seeing a solution* presupposes *seeing a problem*. Although this may sound quite trivial, it is less so if we try to formulate a theory of scientific discoveries. In this case, namely, we have to clarify what it means "to see a problem," and we have to develop a description of scientific activity in a way that the possibility of scientific discoveries can be explained within the framework of this description. Just this is the goal of this paper: to sketch a general description of the scientific method that allows an explanation of creativity in scientific discoveries. My starting point is a terminology that has been developed for the first time by Charles S. Peirce a century ago. The concepts I am mainly focusing on are "abduction" on the one hand, and "diagrammatic reasoning" on the other.

Peirce's famous concept of "abduction" had already an impressive career in philosophy of science as a means for a better understanding of the "context of discovery" (see e.g., Hanson 1972 <1958>; Simon 1979; Nickles 1980b; Nickles 1980a; Grmek et al. 1981; Jason 1988; Kleiner 1993;

Haaparanta 1993; Meheus & Nickles 1999; Magnani et al. 1999; Magnani 2001; Magnani & Nersessian 2002). However, as I will show in the first part of this paper, many things Peirce himself says about abduction are more confusing than helpful when it comes to *explain how* "the process of forming an explanatory hypothesis," as he defines abduction at one point (Peirce CP 5.171), might be possible.

To overcome the shortcomings of Peirce's concept of abduction with regard to explaining the possibility of scientific creativity, I will show in the second part of this paper that it can be argued more convincingly for the relevance of Peirce's concept of abduction when we combine it with his discussion of some other concepts he developed for the first time: "diagrammatic reasoning," "hypostatic abstraction," "theorematic deduction," and the "theoric transformation" of a problem. These concepts are far less known than "abduction," but more interesting, as I will show.

While Peirce introduced abduction besides deduction and induction as one of "three elementary kinds of reasoning" (CP 8.209) that are relevant for any sort of scientific inquiry (cf. CP 7.202–207; CP 6.469–476), his discussion of the other concepts mentioned above is limited to his philosophy of mathematics, or better: of creativity in mathematics. There are hardly any passages in Peirce's writings where he combines both these discussions (cf. Hoffmann 2005a, pp. 203ff.). Such a combination, however, should be interesting for an understanding of creativity in science. Therefore, I will use the terms Peirce introduced in his philosophy of mathematics for a general theory of scientific discovery.

With regard to that, the second part of this paper starts with the concept of "diagrammatic reasoning" and shows then which role the other concepts can play within this activity. Peirce defines "diagrammatic reasoning" as a three-step process: (a) constructing a representation, (b) experimenting with it, and (c) observing the results (Peirce NEM IV 47f.; cf. Stjernfelt 2000; Hoffmann 2003; Hoffmann 2004). The basic idea is according to my interpretation that by representing a problem in a diagram we can not only "play" with this problem, but first of all with our own cognitive and representational means to get access to this problem (cf. for an example in conflict

management Hoffmann 2005c). We have to *represent* what we know—or think to know—in order to *see*, first, its limitations and, second, new possibilities. This latter, creative step will be the place where abduction and the concepts "hypostatic abstraction," "theorematic deduction," and "theoric transformation" enter the stage. Before I come to abduction, I would like to define the other concepts mentioned here because we can use them for the whole discussion.

"Hypostatic abstraction" can be defined as creating a new sign for a new object by transforming a concrete predicate into an abstract noun. Peirce gives a nice example from Moliere's *Malade Imaginaire* where a candidate for a medical degree answers the examination question "why opium puts people to sleep, by saying that it is because it has a dormative virtue" (CP 4.234, 4.463). Although this answers seems to absolutely ridiculous since instead

> of an explanation he simply transforms the premise by the introduction of an abstraction, an abstract noun in place of a concrete predicate. It is a poignant satire, because everybody is supposed to know well enough that this transformation from a concrete predicate to an abstract noun in an oblique case, is a mere transformation of language that leaves the thought absolutely untouched. I knew this as well as everybody else until I had arrived at that point in my analysis of the reasoning of mathematics where I found that this despised juggle of abstraction is an essential part of almost every really helpful step in mathematics. (Peirce NEM IV 160)

"Hypostatization"—from the Greek ὑπόστασις—or "*reification*" (in Latin) means creating a new thing out of what is not a thing. For Peirce it is creating an entity based on a specific sort of abstraction. While in "prescissive abstraction" we abstract from features like color and width to define a geometrical line, in "hypostatic abstraction" we turn what can be a predicate of many things—honey is sweet, strawberries are sweet, sugar is sweet—into "a subject of thought" (CP 5.534): "sweetness" (CP 4.235). Peirce highlights especially "the importance of this operation in mathematics."

> (I)t will suffice to remember that a *collection* is an hypostatic abstraction, or *ens rationis*, that *multitude* is the hypostatic abstraction derived from a predicate of a collection, and that a *cardinal number* is an abstraction attached to a multi-

> tude. So an *ordinal number* is an abstraction attached to a *place*, which in its turn is a hypostatic abstraction from a relative character of a unit of a *series*, itself an abstraction again. (CP 5.534)

"Hypostatic abstraction" can either be the *process* of generating new signs that signify new objects, or the *product* of this process. In general, I would assume that all concepts of our languages are outcomes of hypostatic abstraction performed at some time in the history of our cultures. This way, hypostatic abstraction would be one of the most fundamental concepts to describe the genesis of knowledge.

Hypostatic abstraction is also central for the next concept I mentioned above: "theorematic deduction." Peirce introduced this concept as part of a discussion that treats deductive reasoning—surprisingly—as a creative activity (cf. Hoffmann 2005b, chap. 6). He puts "theorematic deduction" in contrast to "corollarial deduction" as two forms of "necessary deductions."

> A Corollarial Deduction is one which represents the conditions of the conclusion in a diagram and finds from the observation of this diagram, as it is, the truth of the conclusion. A Theorematic Deduction is one which, having represented the conditions of the conclusion in a diagram, performs an ingenious experiment upon the diagram, and by the observation of the diagram, so modified, ascertains the truth of the conclusion. (Peirce CP 2.267; cf. CP 7.204)

The relevance of theorematic deduction becomes again obvious in mathematics. Each time when we perform a proof "by the introduction of auxiliary individuals into the argument" (Hintikka 1983 <1980>, 113, cf. 109f.), we perform a theorematic deduction. That might be a subsidiary line in geometry, or a lemma "when it comes to proving a major theorem" (Peirce EP II 96, 1901). In Peirce's terminology, the introduced "auxiliary individual" would be a hypostatic abstraction; and this hypostatic abstraction could either be newly created or be transferred from another context. Peirce calls the first case an "abstractional" theorematic deduction, and in the second a "non-abstractional" one (NEM IV 49; cf. Hoffmann 2005d), but it seems to be more convincing to call the first one a "creative theorematic deduction" and the second one an "analogical theorematic deduction" since we take

over a hypostatic abstraction that has been created for another purpose. Introducing the term "fission," for example, into a deductive argument in nuclear physics would be a case of "analogical theorematic deduction" since "fission" has already been used earlier in biology to describe "the division of a cell or organism into new cells or organisms, as a mode of reproduction," and in astronomy as the "breaking up of one star into two others, as postulated in one theory of the origin of binary stars" (Oxford English Dictionary).

The last Peircean term I will discuss here as interesting to describe scientific discoveries is "theoric transformation." In spite of the word's similarity to "theorematic," it is very different. While "theorematic" seems to be coined from "theorem," "theoric" refers for Peirce to the Greek "*θεορία*" (our "theory," original meaning: "vision"). He translates this term as "the power of looking at facts from a novel point of view" (Peirce MS 318: CSP 50 = ISP 42). "Theoric" reasoning consists "in the transformation of the problem,—or its statement,—due to viewing it from another point of view" (ibid., CSP 68 = ISP 225). Thus, a "theoric transformation," or a "theoric step" in a deductive argument, means changing the perspective. We are looking at the same data, or the same representation, but in a way that opens up completely new horizons of interpretation. Peirce hints at the well-known fact that especially developing the *idea* of a proof in mathematics often depends on a "theoric" shift (CP 4.612). At several points he uses the proof of Desargues' theorem in projective geometry as an example of how a theoric transformation works (cf. my analysis in Hoffmann 2005a, pp. 170-186; Hoffmann 2005d).

Abduction

The problems we are facing with Peirce's concept of abduction can be illustrated when we simply take a look at one his best known definitions of this term (I am referring here only to what he developed after 1900 since many of his earlier considerations are insufficient, as he admits for himself: Peirce L 409, ISP 73; CP 2.102, 8.227; cf. Hoffmann 2005b, pp. 187ff.):

> Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis. (Peirce CP 5.171)

The second half of this quote is not part of the definition, but an explanation for it. However, it adds something to this definition because it says implicitly that there are *only* three logical operations for Peirce, a claim that he confirms in another remark where he says that "there are but three elementary kinds of reasoning": abduction, deduction, and induction (CP 8.209). This means, however, that any form of "reasoning," or "logical operation," that is neither deduction nor induction has to be *abduction*; and that might be much more than we would expect at a first glance. According to Peirce, we find abduction not only in science as the process of "examining a mass of facts and in allowing these facts to suggest a theory" (CP 8.209), but also in any *perception* "when I so much as express in a sentence anything I see" (Peirce LOS, p. 899f.; cf. CP 5.182ff., 8.64). Even when "a chicken first emerges from the shell" and "does not try fifty random ways of appeasing its hunger, but within five minutes is picking up food, choosing as it picks, and picking what it aims to pick," this is "just like abductive inference" (Peirce LOS, p. 899f.). This unspecified broadness of possibilities to apply the concept of abduction has led to many attempts in the literature to develop classifications of different *forms* of abduction (e.g., Bonfantini & Proni 1983; Eco 1983; Shank & Cunningham 1996; Magnani 2001).

A second problem concerns the claim that abduction is supposed here to be a "logical operation" (cf. Kapitan 1992). Elsewhere I showed that this claim can only be convincing if we start from one of two possibilities: either from a very broad understanding of "logic," for example "logic" defined as "the *art of devising methods of research*,—the *method of methods*" (CP 7.59; cf. 3.618, 4.227, Fann 1970, p. 23f.); or we interpret "logical operation" as referring only to the *form* Peirce uses to describe abductive reasoning, but not to the creative act of "forming an explanatory hypothesis" (cf. Hoffmann 1999). This second possibility is evident when we look at the famous "perfectly definite logical form" of abduction the late Peirce defined as follows:

(P1) "The surprising fact, C, is observed;"
(P2) "But if A were true, C would be a matter of course,"
(C) "Hence, there is reason to suspect that A is true" (CP 5.189).

Peirce himself states in the next sentence that according to this form "A cannot be abductively inferred ... until its entire content is already present" in the second premise. That means, however, that the logical form for itself leaves the question unanswered how to get the hypothesis "A." There is no way to create a new hypothesis by a *logical* inference. "The first emergence of this new element into consciousness must be regarded as a perceptive judgment" (CP 5.192). Only *after* the new hypothesis has emerged, we can connect "this perception with other elements" (ibid.) in a logical form like the one quoted above.

However, if we accept this distinction between "*inferential aspects of abduction*" and the creative, "*perceptive aspects of abduction*" (Hoffmann 1999, p. 280), then it is hard to see how abduction can be the "the only *logical* operation which introduces any new idea" as Peirce claims in our initial quote (my emphasis). This claim could be justified only if "logical operation" means simply "method," or "strategy" (cf. Paavola & Hakkarainen 2005). Otherwise we would get a contradiction to what we discussed just a moment ago.

The third problem of this famous quote concerns the question what exactly abduction is supposed to do. In the quote, Peirce hints, on the one hand, at the "the process of forming an explanatory hypothesis" and, on the other, at introducing a "new idea." This, however, allows different interpretations. First, it is possible that the two concepts refer to two different operations, because we can form an "explanatory hypothesis" without generating a "new idea." If any *perception* is a case of abduction, we would generate a lot of "explanatory hypotheses" regarding sensory inputs though hardly any "new idea." For example, when reading a word, the word we read is a hypothesis that "explains" a sequence of letters. In this case, we form an explanatory hypothesis without introducing a new idea since the idea we associate with a certain sequence of letters exists already in our mind.

Second, it is not clear whether the "idea" we introduce is only new for us as individuals or new for our civilization. And third, an "idea" can either be what we discussed above as a "hypostatic abstraction," that is something that can be represented by a singular concept, or by a symbol, or it could be an new *perspective* on the same data as produced by a "theoric transformation." In sum, Peirce's innocent and well-known definition can describe *six very different forms of abduction* (cf. Table 1).

	If "idea" means something that can be represented by a singular symbol (hypostatic abstraction)	If "idea" means the perspective on data, or on a representation (theoric transformation)
If an explanation is possible by referring to an idea that exists already in our mind	identifying abduction	Gestalt shift
If we create an idea that is new for us, but that exists already as a part of our culture's knowledge	analogical hypostatizing abduction	analogical theoric abduction
If we create an idea that is entirely new	creative hypostatizing abduction	creative theoric abduction

Table 1: Six forms of abduction that are possible based on CP 5.171 alone

At this point we can already see that by means of the concepts Peirce developed in his philosophy of mathematics we can specify some important distinctions that might be overlooked without this terminology. The four creative forms of abduction listed in the second and third row of this table will be very useful later on.

Peirce himself does not specify different forms of abduction precisely enough. However, what is important is his description of the general *situation* in which abduction takes place. This situation can be determined by two points: (a) abduction starts from the *particular* as it is given in perception, and (b) it leads to something *general*. This general might be, first, a hypothesis, or a theory that is "needed to explain ... surprising facts" (CP 7.218) or, second, "a statement of fact" by which we "make intelligible" any image provided by perception (LOS 899) or, third, an activity that is driven by a general habit,

or an "instinct," like the behavior of the chicken mentioned above that from early on picks the right things (ibid.).

Two things are important to note: on the one hand, that this general itself is, as Peirce once said, "in no wise contained in the data from which it sets out"; it is "entirely foreign to the data" (LOS, 898f.); and, on the other hand, that "there is no force" in this process leading from the particular to the general (CP 8.209), meaning that—maybe except in cases of perception and instinct-driven behavior—it is not predetermined what the general exactly will be by which we react to a certain particular event or situation.

Having no force on the one hand, but being more successful than could be expected by pure chance on the other (CP 5.172, 5.591), leads to the question how our ability to infer abductively can be explained. Peirce himself hints at many places at an "instinctive" power of "guessing rightly" (cf. Rescher 1995; Fann 1970, pp. 35–38; Paavola & Hakkarainen 2005), and also at "the uncontrolled part of the mind" (CP 5.194; cf. Semetsky 2005). But since none of these approaches is sufficiently elaborated, the possibility of abductive reasoning remains at the end unexplained. In this situation, we should focus on *methods* that might, at least, facilitate abduction. Just this is the role "diagrammatic reasoning" can play in a general theory of scientific discoveries.

Diagrammatic reasoning

The essential feature of diagrammatic reasoning that makes this method so interesting for the description of scientific discoveries is not the three-step process I mentioned above as its definition: constructing a diagram, experimenting with this diagram, and observing the results (Peirce NEM IV 47f.). The essential point is rather that the whole process has to be performed by means—and within the limits—of a given "system of representation" (CP 4.418). When Peirce defined a "diagram" as "a representamen which is predominantly an icon of relations" that "should be carried out upon a perfectly consistent system of representation" (CP 4.418), he worked on a series of definitions to formulate the foundations of his so-called "Existential Graphs," a graphical notation of logic intended to replace

algebraic notations (cf. Roberts 1973; Ketner 1996 <1990>; Shin 2002). This system of representation is characterized by a set of conventions to represent propositions and logical relations between those propositions, and a set of rules for the transformation of graphs. From a logical point of view, Peirce's Existential Graphs—at least the Alpha and Beta Part—are as sound and complete as symbolic systems of logic.

There is no question that one needs "a perfectly consistent system of representation" for representing logical implications and validity, and the same is true for representation systems in mathematics. (Although we know now that there is a principle limit of consistency *and* completeness imposed on us by Gödel's incompleteness theorems.) If we want to prove in Euclidean geometry, for example, that the triangle's inner angles sum up to 180°, we can use a parallel to the triangle's base as auxiliary line to perform the proof. The possibility of such a parallel, however, is provided only in Euclidean geometry, not in non-Euclidean ones. That is, the specific means available in this system of representation *determine* the outcome of any transformation of a Euclidean figure we perform in accordance with the rules of this system. At the same time, however, those operations are also *constrained* by the means available. If we change the system of representation—as has been done in the development of non-Euclidean geometries—we open up entirely new possibilities (cf. Hoffmann 2004). In short, a proof in mathematics is only as perfect and consistent as the representation system in which it is performed.

The consistency of representation systems is also essential when we use the concept of diagrammatic reasoning—beyond the limits of logic and mathematics—for a general theory of scientific discoveries. Not only axiomatic systems in mathematics have to be consistent, but also, for instance, the description of styles in art, the grammar we construct to understand the syntax of our everyday languages, and theories in science.

Besides consistency, there are two further features that are common to all those systems of representation. On the one hand, we need them to design a particular representation. We need an axiomatic system to construct a proof in mathematics; we need a scientific theory to formulate a hypothesis

or observational statement because otherwise the meaning of the concepts we are using and the inferences we are performing could not be justified; we need the grammar of our everyday language to formulate a normal sentence; and we need a style to draw a painting in art—although in both the latter cases knowledge about the system of representation might be mostly implicit. On the other hand, all these systems of representation play a *normative* role: in logic and mathematics we can check the validity of an inference, or a proof, by means of the rules and conventions defined by a certain system of representation; in science the plausibility of a particular statement depends on its theoretical background; in our everyday languages the correctness of expressions is determined by grammar; and in art the definition of styles is a means of classification, for example.

The central role of the chosen representation system for diagrammatic reasoning becomes visible when we consider its function for each of the three steps by which diagrammatic reasoning can be defined (cf. Figure 1).

As this "map definition" of diagrammatic reasoning shows, the outcome of each of the three steps depends—in different ways—on the system of representation we choose in a certain situation to represent a problem, or the knowledge area we are focusing on.

The consistency and normativity of the chosen system of representation are decisive when it comes to the possibility of discoveries. Peirce once highlighted as a core idea of this *pragmatism* that all reasonings—and especially mathematical reasonings—

> turn upon the idea that if one exerts certain kinds of volition, one will undergo in return certain compulsory perceptions. Now this sort of consideration, namely, that certain lines of conduct will entail certain kinds of inevitable experiences is what is called a "practical consideration." (CP 5.9)

Such an "inevitableness" depends obviously on the normative character of the representational system in which such reasoning is performed, as is evident from mathematics: That *2* plus *2* equals *4* results from the rules and conventions of arithmetic as the chosen system of representation.

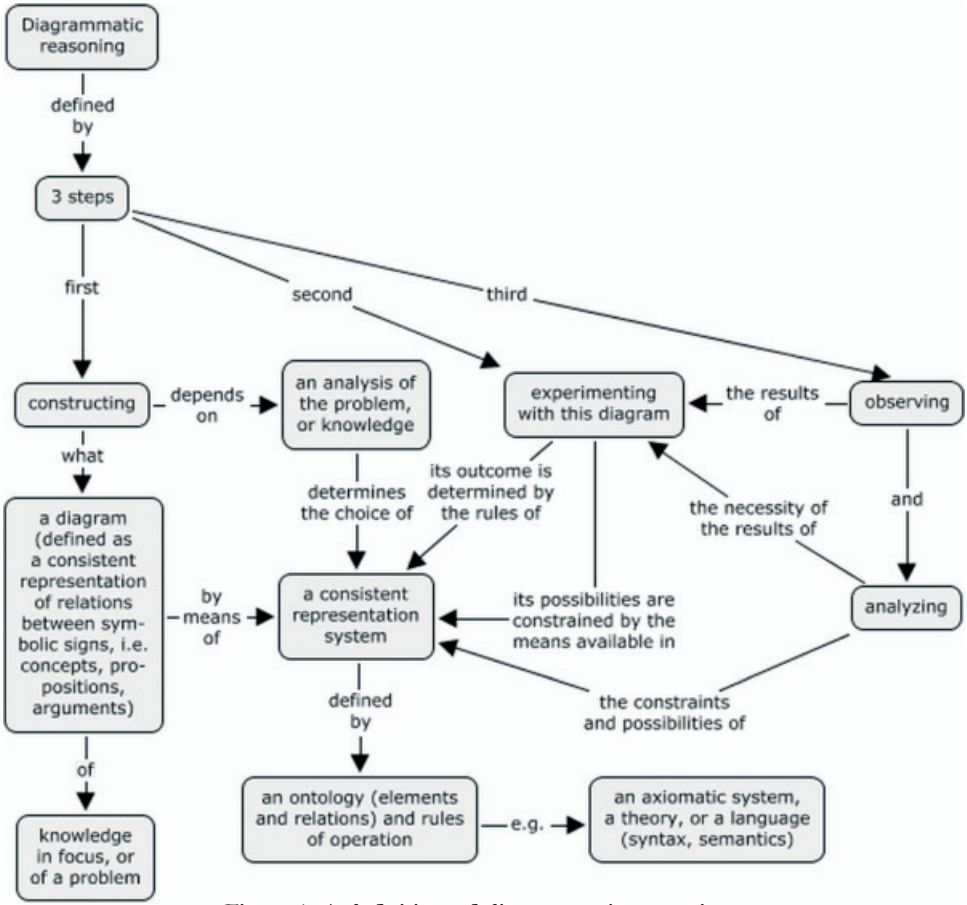


Figure 1: A definition of diagrammatic reasoning.
All maps are created with IHMC Cmap tools: http://cmap.ihmc.us/

This "inevitable experience" resulting from rule-driven activity is the most important precondition of discovering something *new* by diagrammatic reasoning. As I showed elsewhere with regard to mathematics, we can distinguish two possibilities: on the one hand, the process of unfolding new *implications* of a diagram within a *given* system of representation and, on the other, the process of *developing* representational systems *themselves* that can open up new horizons and possibilities—as it has been the case in the development of non-Euclidean geometries (Hoffmann 2004).

The first case refers to the fact that we never can have a complete overview of all the implications of what we know already. Only experimentation with representations in concrete situations reveals what might already be given implicitly in our own systems of knowledge. Peirce described this case of discovering something new by saying that a diagram constructed by a mathematician "puts before him an icon by the observation of which he detects relations between the parts of the diagram other than those which were used in its construction" (NEM III 749). By experimenting upon the diagram and by observing the results thereof, it is possible, as he says, "to discover unnoticed and hidden relations among the parts" (CP 3.363).

The second case is closely connected with the problem of consistency. It might be that a rule-driven experimentation with diagrams brings to light inconsistencies or undecidable situations within our chosen system of representation. If we don't have any reason to doubt the correctness of the diagrammatic transformations we have performed, we are forced to improve the system of representation we have used.

In this situation, the *compelling* character of diagrams and the "inevitable experience" we make in diagrammatic reasoning are decisive. The results of experiments have to "stand up against our consciousness," as Hull (1994, p. 282) puts it, because only in this case can a diagram "compel us to think quite differently" (CP 1.324). It is only if we have already certain expectations concerning what should happen in processes of diagrammatization that there is a need to develop something new when those expectations are frustrated.

For a *general* theory of scientific discovery which goes beyond the limits of logic and mathematics, we can start from a critique Peirce formulated once with regard to Kant. Kant saw already that in mathematics the drawing of "necessary consequences" is possible because "the mathematician uses what, in geometry, is called a 'construction,' or in general a diagram, or visual array of characters or lines" (CP 3.560). But while for Kant the use of constructions is the criterion to distinguish been mathematics and philosophy (Kant CPR B 741), Peirce insists that there is no difference between both when it comes to the central role of "constructions," or "diagrams":

> All necessary reasoning whatsoever proceeds by constructions; and the only difference between mathematical and philosophical necessary deductions is that the latter are so excessively simple that the construction attracts no attention and is overlooked. (CP 3.560; cf.. 5.147f.)

> All necessary reasoning without exception is diagrammatic. That is, we construct an icon of our hypothetical state of things and proceed to observe it. This observation leads us to suspect that something is true, which we may or may not be able to formulate with precision, and we proceed to inquire whether it is true or not. (CP 5.162)

The connection between "necessary reasoning" and "diagrammatic reasoning" results from the normative character of the used representation system, as we saw above. However, if we try to generalize what is plausible with regard to mathematics, we are facing a serious problem. While it is convincing to emphasize the role of *representations* in any science, it is harder to determine what exactly the "systems of representation" in science are that we need to explain the possibility of scientific discoveries by means of diagrammatic reasoning. While in mathematics and logic we can mostly formulate a clear distinction between a "consistent system of representation" (e.g., an axiomatic system, or a notation and conventions) and a "diagram" constructed by means of this system, this is not so easy in the sciences. Of course, we could say that any science uses different—more or less formal—languages so that its representations depend on the "grammar" of those languages. But just this is part of the problem. There are quite a lot of those "languages," and they are more or less relevant for what happens in scientific activity. We would assume, for example, that it does not matter whether we formulate Einstein's theories of relativity in English or in Chinese. Sometimes, however, it could matter, for example when we formulate a sociological theory that is based on ideas, or concepts, for which there is no equivalent in another culture. And it matters of course that we need for Einstein's theories the language of mathematics; sometimes mathematical theories must be developed before there is progress in physics.

The problem is, thus, that there are very different systems of representation that we often use at the same time when representing scientific knowl-

edge, and that not all of them determine and constrain the possibilities of scientific activities in the same way as the Euclidean language determines proof possibilities in geometry. It seems to be impossible to propose a classification of all systems of representation we are using in different disciplines, and to determine in general which is relevant in which way. The problems we had to face for such an endeavour would be similar to the problems Thomas S. Kuhn faced when forced to clarify his concept of "paradigm" which is—at least with regard to its normative character—pretty similar to our concept of "representation system."

The only way I can see to cope with this problem is not to start with a "top down" classification of representation systems, but to proceed "bottom up." It should be sufficient to define the term "system of representation" only formally as that consistent set of conditions we have to presuppose to *understand* a representation, and to look at diagrammatic representations themselves guided by the question 'What kind of representation system do we need to make sense out of this expression?'. When we read a scientific text in which Sun, Moon, and Mars are called "planets," for example, a consistent representation system that could produce such an utterance would be the Ptolemaic astronomy, but not the Copernican system of representation with its classification of celestial bodies (Chen et al. 1998, p. 9).

From a general point of view, the term "system of representation" refers first of all to those conditions of understanding. Therefore, those systems can only be determined based on an analysis of concrete representations, they are *relative* to what they are supposed to determine. It depends on the concrete diagram what we have to identify as a relevant system of representation. Relevant is what we need to understand the diagram.

In order to show now how based on these considerations the possibility of scientific discoveries can be explained, I suggest to extend the definition of diagrammatic reasoning proposed in Figure 1 above. We have to continue the series of steps that are described in this definition. A complete description of what follows after the third step described in Figure 1 is presented in Figure 2. Remember, the first three steps that are already mentioned by Peirce are: constructing a diagram, experimenting with it, and observing—

and analyzing—the outcome of those experiments. The continuation of the process depends, obviously, on the result of the third step. As elaborated in Figure 2, we can distinguish three possible results, each combined with a certain implication that will then again lead—at least in two cases—to further steps.

The basic ideas of this enlarged model of diagrammatic reasoning can be summarized as follows:

1. There is only one situation in which the process of diagrammatic reasoning comes—temporarily—to a halt: when we get the first possible result of observing and analyzing the outcome of an experiment, that is "the outcome does not contradict our expectations" (see Figure 2). In this situation we see that what we observe in the experiment is a necessary implication of the original diagram, and we can be happy that no contradiction occurs. There is no reason to continue the process. We learned something new, although nothing that could surprise us. It is like proving that the triangle's inner angles sum up to 180° when nobody thought about that before. It is interesting, but no reason to doubt the reliability of our cognitive and representational means.
2. Everything else besides this first possibility leads over a series of further steps finally back again to the first step, the construction of a diagram. This means, the whole process is with regard to the *activities* described in Figures 1 and 2 an endless loop that stops only when the situation mentioned in (1.) occurs.
3. However, regarding the *representational means* available, each of these loops continues with the first step on a more advanced level: we are either able to construct a *new diagram* by the means of the representation system that we used already from the very beginning, or we are able to construct a new diagram based on a *new system of representation*.
4. The possibility of such an advancement is the result of scientific discoveries that are differentiated in this model as eight forms of abduction (1. to 8. in the second half of Figure 2). The terminol-

ogy according to which these forms are classified goes back to the Peircean concepts "hypostatic abstraction" and "theoric transformation" as defined at the end of the introduction, and to the terms I coined based on that with regard to abduction in Table 1. Each of these forms of abduction enlarge the set of *representational*—and therefore also: *cognitive* (cf. Hoffmann & Roth forthcoming)—means available for diagrammatic reasoning.

5. The distinction of all the possibilities differentiated in Figure 2 is based on an analytical interest to separate different forms of scientific discoveries—or elements of those discoveries—as clear as possible. In practice, however, these forms will usually be intertwined, I guess.

This leads us back again to our starting point, the historical example I discussed in the beginning. It is interesting to see that the majority of physicists' "blindness" regarding the possibility of nuclear fission in 1934 can directly be ascribed to the limitations of the predominantly used "system of representation." The "diagrams and notations which were developed," says Anderson with reference to some diagrams published by Meitner in 1934, "could only represent the idea that a projectile hits a nucleus which as a result transformed into another nucleus by the emission of a particle" (Andersen 2006, p. 7). The majority of physicists was convinced by the obvious harmony between the taxonomy of disintegration processes mentioned above and "Gamow's theory of decay which precluded decay products larger than the α-particle" (ibid.). The thesis "that only particles up to the size of the α-particle could be emitted would become tacitly accepted in the whole scientific community to such an extend that the mere possibility of larger decay products [and, therefore, also the possibility of nuclear fission, M.H.] would never be mentioned" (ibid.). Background assumptions like this one form a "system of representation" in so far as they can be identified as those *conditions* that make concrete "diagrams," that is any scientific utterance formulated in this context, *understandable*.

The idea of *splitting* the nucleus into larger fragments was simply beyond the possibilities of those representational means, and since hardly anybody

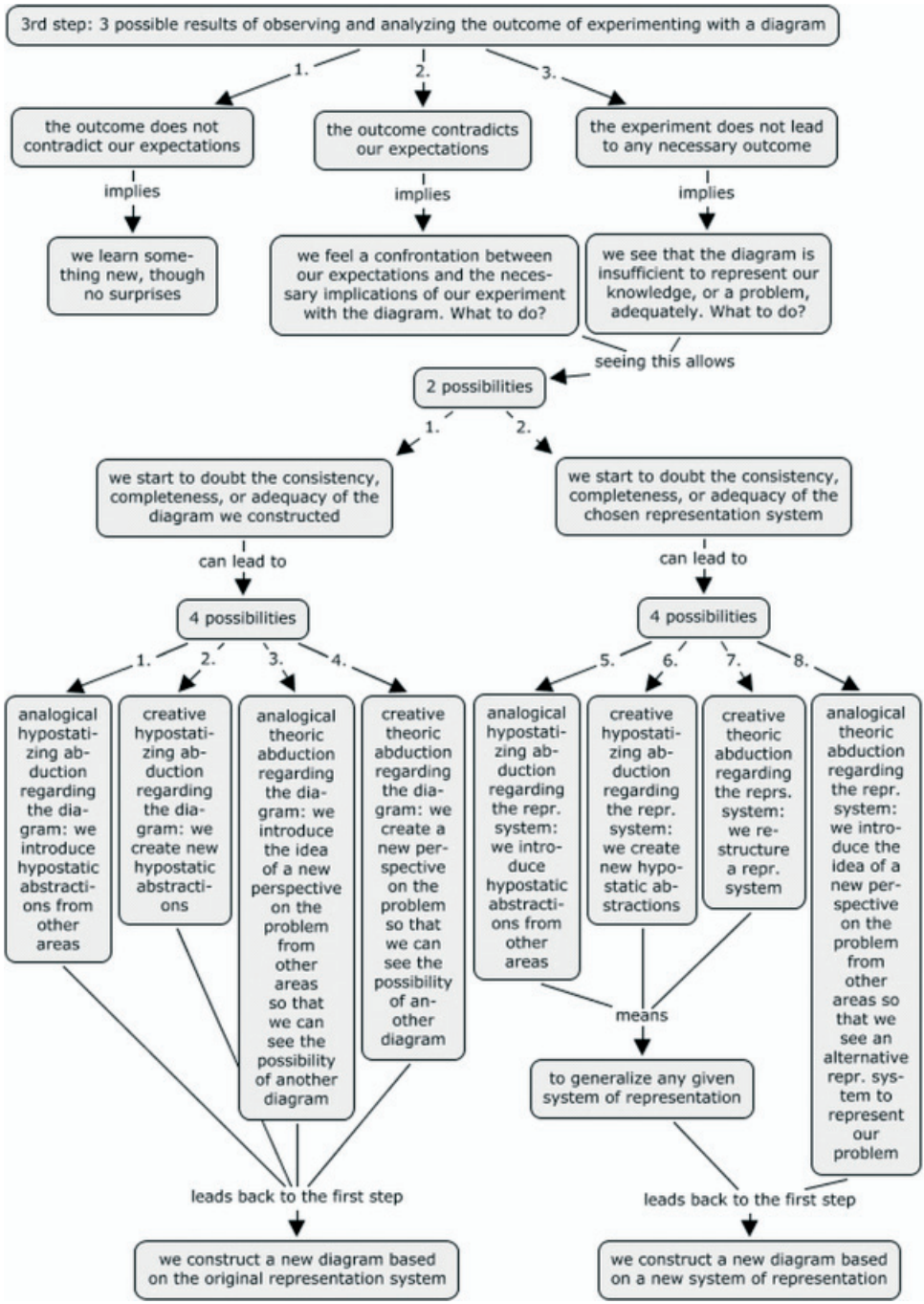


Figure 2: The process of diagrammatic reasoning, extended version, continuing the series of three steps described in Figure 1.

saw a need to doubt this well-established system of representation, there was no room for a discovery as revolutionary as that of nuclear fission. Seeing a solution presupposes seeing a problem. This way, the majority of scientists was happy with what is described in Figure 2 as the first possible result of observing the outcome of experiments with diagrammatic representations. They learned something new through Fermi's experiments, but no big surprises.

For Ida Noddack, on the other side, the broadly accepted assumption of Fermi's that he produced in his experiments a transuranic element contradicted directly her expectations. These expectations, however, were based on a *different* system of representation: not the taxonomy of "disintegration processes" that fitted so well to Gamow's theory, but a system of representation that she developed over the years when working on the chemical properties of elements that were yet to discover. In this situation of *seeing* a severe problem regarding the consistency between her predominant system of representation and what Fermi described, she did not doubt this system, but the assumption of Fermi's. To formulate it in the language of my extended model of diagrammatic reasoning (cf. Figure 2), she started "to doubt the consistency, completeness, or adequacy" of a possible diagram that—constructed by the means of her own system of representation—should be able to include Fermi's thesis. In this situation, she performed what can be described as the third form of abduction counted in Figure 2: an "analogical theoric abduction." Since she writes in her paper that heavy nuclei might "decay" when bombarded with neutrons ("Es wäre denkbar, daß bei der Beschießung schwerer Kerne mit Neutronen diese Kerne in mehrere größere Bruchstücke zerfallen, die zwar Isotope bekannter Elemente, aber nicht Nachbarn der bestrahlten Elemente sind"; Noddack 1934), it is clear that she looks at the experiment—compared to Fermi—from "a novel point of view," that is she performs what Peirce calls a "theoric transformation." And since the idea of "decay" occurs in many areas of experience, it is not a "creative theoric abduction," but an "analogical" one. If she had already coined a new term like "Kernspaltung" (nuclear fission) for what she hy-

pothesized based on this analogical theoric abduction, she would have performed additionally a form of "hypostatizing" abduction.

The essential point is, however, that each of these eight possible forms of abduction leads to an enlargement of the representational and cognitive means that are then available for further processes of diagrammatic reasoning. And this again is what scientific progress is all about.

Acknowledgement

Many thanks to Professor Olga Pombo and Alexander Gerner for organizing a wonderful workshop in Lisbon in May 2006, and for inviting me to present my research.

References

Andersen, H. (2006). Unexpected Discoveries, Graded Structures, and the Difference between Acceptance and Neglect. In J. Meheus & T. Nickles (Eds.), *Models of Discovery and Creativity* (pp. 1-27): Springer.

Bonfantini, M. A., & Proni, G. (1983). To Guess or Not To Guess? In U. Eco & T. A. Sebeok (Eds.), *The Sign of Three. Dupin, Holmes, Peirce* (pp. 119-134). Bloomington: Indiana Univ. Press.

Chen, X., Andersen, H., & Barker, P. (1998). Kuhn's theory of scientific revolutions and cognitive psychology. *Philosophical Psychology 11*, 5-28.

Eco, U. (1983). Horn, Hooves, Insteps. Some Hypotheses on Three Types of Abduction. In U. Eco & T. A. Sebeok (Eds.), *The Sign of Three. Dupin, Holmes, Peirce* (pp. 198-220). Bloomington: Indiana Univ. Press.

Fann, K. T. (1970). *Peirce's Theory of Abduction*. The Hague: Nijhoff.

Grmek, M. D., Cohen, R. S., & Cimino, G. (Eds.). (1981). *On Scientific Discovery: The Erice Lectures 1977*. Dordrecht Reidel.

Haaparanta, L. (1993). Peirce and the Logic of Logical Discovery. In E. C. Moore (Ed.), *Charles S. Peirce and the Philosophy of Science. Papers from the Harvard Sesquicentennial Congress* (pp. 105–118). Tuscaloosa and London: The University of Alabama Press.

Hanson, N. R. (1972 <1958>). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambride University Press.

Hintikka, J. (1983 <1980>). C.S. Peirce's 'First Real Discovery' and its Contemporary Relevance. In E. Freeman (Ed.), *The Relevance of Charles Peirce* (pp. 107–118). La Salle, Ill.: Hegeler Institut.

Hoffmann, M. H. G. (1999). Problems with Peirce's Concept of Abduction. *Foundations of Science, 4*(3), 271–305.

Hoffmann, M. H. G. (2003). Peirce's "Diagrammatic Reasoning" as a Solution of the Learning Paradox. In G. Debrock (Ed.), *Process Pragmatism: Essays on a Quiet Philosophical Revolution* (pp. 121-143). Amsterdam: Rodopi.

Hoffmann, M. H. G. (2004). How to Get It. Diagrammatic Reasoning as a Tool of Knowledge Development and its Pragmatic Dimension. *Foundations of Science, 9*(3), 285-305.

Hoffmann, M. H. G. (2005a). *Erkenntnisentwicklung. Ein semiotisch-pragmatischer Ansatz.* Frankfurt am Main: Klostermann.

Hoffmann, M. H. G. (2005b). *Erkenntnisentwicklung. Ein semiotisch-pragmatischer Ansatz (Knowledge development. A semiotic and pragmatic approach).* Frankfurt am Main: Klostermann.

Hoffmann, M. H. G. (2005c). Logical argument mapping: A method for overcoming cognitive problems of conflict management. *International Journal of Conflict Management, 16*(4), 305–335.

Hoffmann, M. H. G. (2005d). Signs as means for discoveries. Peirce and his concepts of "Diagrammatic Reasoning," "Theorematic Deduction," "Hypostatic Abstraction," and "Theoric Transformation". In M. H. G. Hoffmann, J. Lenhard & F. Seeger (Eds.), *Activity and Sign - Grounding Mathematics Education* (pp. 45-56). New York: Springer.

Hoffmann, M. H. G., & Roth, W.-M. (forthcoming). The complementarity of a representational and an epistemological function of signs in scientific activity. *Semiotica.*

Hull, K. (1994). Why Hanker After Logic? Mathematical Imagination, Creativity and Perception in Peirce's Systematic Philosophy. *Transactions of the Charles S. Peirce Society, 30*, 271–295.

Jason, G. (1988). *The Logic of Scientific Discovery*. New York: P. Lang.

Kant, I. (CPR). *Critique of pure reason* (P. Guyer & A. W. Wood, Trans.). Cambridge 1998: Cambridge Univ. Pr.

Kapitan, T. (1992). Peirce and the Autonomy of Abductive Reasoning. *Erkenntnis, 37*, 1–26.

Ketner, K. L. (1996 <1990>). *Elements of Logic. An Introduction to Peirce's Existential Graphs* (2. ed.). Lubbock: Arisbe Associates.

Kleiner, S. A. (1993). *The Logic of Discovery: A Theory of the Rationality of Scientific Research*. Dordrecht Kluwer.

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

Magnani, L., & Nersessian, N. J. (Eds.). (2002). *Model-Based Reasoning: Science, Technology, Values*. New York: Kluwer Academic / Plenum Publisher.

Magnani, L., Nersessian, N. J., & Thagard, P. (Eds.). (1999). *Model-based Reasoning in Scientific Discovery*. New York: Plenum Publishers.

Meheus, J., & Nickles, T. (Eds.). (1999). *Discovery and Creativity*. Special issues of the Journal: Foundations of Science 4(3) and 4(4).

Nickles, T. (1980a). Introduction. In T. Nickles (Ed.), *Scientific Discovery: Case Studies* (pp. 1–62). Dordrecht: Reidel.

Nickles, T. (Ed.). (1980b). *Scientific Discovery, Logic and Rationality*. Dordrecht: Reidel.

Noddack, I. (1934). Über das Element 93. *Angewandte Chemie, 47*, 653–655.

Paavola, S., & Hakkarainen, K. (2005). Three Abductive Solutions to the Meno Paradox – with Instinct, Inference, and Distributed Cognition. *Studies in Philosophy and Education, 24*(3–4), 235–253.

Peirce. (CP). *Collected Papers of Charles Sanders Peirce*. Cambridge, Mass.: Harvard UP.

Peirce. (EP). *The Essential Peirce. Selected Philosophical Writings. Vol. 1 (1867-1893), Vol. 2 (1893-1913)*. Bloomington and Indianapolis 1992 +1998: Indiana University Press.

Peirce. (L). *Letter*. Available in the Peirce Microfilm edition. Pagination: CSP = Peirce / ISP = Institute for Studies in Pragmaticism.

Peirce. (LOS). *Historical Perspectives on Peirce's Logic of Science: A History of Science*. Berlin 1986: Mouton.

Peirce. (MS). *The Charles S. Peirce Papers. Manuscript Collection in the Houghton Library, Harvard University*. Available in the Peirce Microfilm edi-

tion. Pagination: CSP = Peirce / ISP = Institute for Studies in Pragmaticism.

Peirce. (NEM). *The New Elements of Mathematics by Charles S. Peirce* (Vol. I-IV). The Hague-Paris/Atlantic Highlands, N.J., 1976: Mouton/Humanities Press.

Rescher, N. (1995). Peirce on Abduction, Plausibility, and the Efficiency of Scientific Inquiry. In N. Rescher (Ed.), *Essays in the History of Philosophy* (pp. 309–326). Aldershot Avebury.

Roberts, D. D. (1973). *The Existential Graphs of Charles S. Peirce*. The Hague Mouton.

Semetsky, I. (2005). Learning by abduction: A geometrical interpretation. *Semiotica, 157*(1/4), 199–212.

Shank, G., & Cunningham, D. J. (1996). Modeling the Six Modes of Peircean Abduction for Educational Purposes. *Online*, http://www.cs.indiana.edu/event/maics96/Proceedings/shank.html (accessed July 30, 2006).

Shin, S.-J. (2002). *The Iconic Logic of Peirce's Graphs*. Cambridge MA: MIT Bradford Books.

Simon, H. A. (Ed.). (1979). *Models of Discovery and Other Topics in the Methods of Science*. Dordrecht: Reidel.

Stjernfelt, F. (2000). Diagrams as Centerpiece of a Peircean Epistemology. *Transactions of the Charles S. Peirce Society, 36*, 357–384.

Alterity and dialogism in abductive processes of scientific discovery

Augusto Ponzio
Dipartimento di Pratiche linguistiche e Analisi di Testi
Università degli Studi di Bari
augustoponzio@libero.it

Premise

In its more innovative aspects at the basis of the abductive process of scientific discovery, abduction ventures beyond the limits of a defined totality without the guarantee of return or reconciliation to the principles that exist in it. There is a movement towards alterity which more than in terms of intentionality (the latter belongs to objectification and the subject/object relation) or of need, we might express as desire: desire of the absolute other.

The category of Originality dominates in abduction, 'Originality is being such as that being is, regardless of aught else' (CP 2.89). It is precisely this capacity of being regardless of anything else that constitutes alterity. The other is other because of its being independently of reference to a viewpoint, a function, an objective, a relation of distinction or opposition, or of insertion into a unitary story. For this reason, the other is a surplus external to the totality, to the totality of the Self and Sameness which in so far as being a unity, a teleological organization, is in the order of binarity and mediation. Firstness, or Orience, or Originality is 'something which is what it is without reference to anything else within it or without it, re gardless of all force and of all reason' (CP 2.85). For this reason it cannot be incorporated by the totality, but on the contrary stimulates a breach, a renewal, the reopening and reorganization of a totality which is never definitively concluded and systematized. All knowledge, totality, binarity and mediation, all cognition

as adjustment to objects, presuppose orience, that is, alterity. The latter being the lack of adjustment par excellence, the surpassing of the objectifying thought, of the subject/object, and means/end relation.

In abduction the relation between premises and conclusion is iconic and dialogic in a substantial sense, in other words, it is characterized by high degrees of dialogism. Abductive inferential processes of scientific discovery engender sign processes at the highest levels of otherness and dialogism.

1. Modeling

Our definition of semiosis or sign process is centered around the notion of interpretant. The interpretant mediates between *objectum* (from *obicio*), that is, solicitation (interpretandum) and response (sign behavior). In Peirce's view, such mediation distinguishes a semiosis from a mere dynamical action — 'or action of brute force' — which takes place between the terms forming a pair. Instead, semiosis results from a triadic relation which involves a cooperation of three subjects, a sign, its object, and its interpretant', and is not 'in any way resolvable into action between pairs' (Peirce CP 5.484, unless otherwise stated, the numbers in brackets in this chapter refer to Collected Papers, by Charles S. Peirce).

The interpretant does not occur in physical phenomena or in nonbiological interactions, but it occurs only in the organic world.

The terms 'model' and 'modeling' are used in the present text as understood by Thomas A. Sebeok and his global semiotics. 'Modeling' is a pivotal notion used in Sebeok's global semiotics to explain life and behaviour among living entities conceived in terms of semiosis. Therefore, global semiotics also involves modeling systems theory.

Modeling is the foundation of communication in all living beings. Communication necessarily occurs within the limits of a world and its characteristics as modeled by a given species, a world that is species-specific. Jakob von Uexküll speaks of invisible worlds to indicate the domain which englobes all animals according to the species they belong to. What an animal perceives, craves, fears and predates is relative to its own world. Human communication is the most complex and varied form of communication in

the sphere of biosemiosis – the human animal is capable of modeling a potentially infinite number of possible worlds. Sebeok develops the concept of modeling from the so-called Moscow-Tartu school, though he enriches it by relating it to the concept of *Umwelt* as formulated by Jakob von Uexküll

Semiosis may be interpreted as the capacity with which all life-forms are endowed to produce and comprehend the species-specific models of their worlds. Primary modeling is the innate capacity for simulative modeling in species-specific ways.

The so-called Moscow-Tartu school limits the concept of modeling to the human sphere (Lotman's 'semiosphere') and distinguishes between the 'primary modeling system,' an expression used to denote natural language, and the 'secondary modeling system,' used for all other human cultural systems. Instead, Sebeok extends the concept of model beyond the domain of anthroposemiosis, and connects it to the concept of *Umwelt* as elaborated by the biologist Jakob von Uexküll, which in Sebeok's interpretation may be translated as 'outside world model.'

On the basis of research in biosemiotics, we now know that the modeling capacity is operative in all life forms. All life forms are endowed with a capacity for semiosis, therefore the capacity to produce and comprehend the species-specific models of their worlds. Primary modeling is the innate capacity of organisms for simulative modeling in species-specific ways.

The primary modeling system of the species *Homo* is what we may call, with Sebeok, *language,* which should not be confused with verbal language, as in the Moscow-Tartu school and in Noam Chomsky.

The Moscow-Tartu school, specifically Lotman, considered verbal language, and consequently any particular natural language, as a primary modelling system.

But a distinction must be made between *language* understood as 'verbal language,' that is, *as a communication system and a secondary modelling system*, and *language* understood *as a species-specific primary modeling device*. All animal species have a specific modelling device. But the species-specific trait of the human being is a modeling device capable of inventing *many* worlds – as Leibniz say infinite possible worlds – differently from other animals. Instead any other

animal species has only one world. Human beings was endowed with this particular modeding device when they first appeared in evolutionary development; and this specific device conditions evolution determing the capacity for articulate speech in Homo sapiens and Sapiens sapiens.

With Peirce, Sebeok calls this species-specific human modeling capacity (a capacity for the 'play of musement') 'language' and distinguishes this from 'speech': originally language was mute. As a modelling device capable of modelling different worlds, it allowed for communication. Therefore, language thus understood is at the basis of the different historical natural languages. This is the 'lingua mutola' (mute language) discussed by Vico: originally all nations were mute and expressed themselves through acts or body language (434).

The capacity for abduction is strictly connected with this species-specific human modelling.

Secondary and tertiary modelling systems presuppose language understood as a modeling device, therefore, these too indicate uniquely human capacities. In Sebeok's terminology, the secondary modelling system is verbal language or, *speech*; while tertiary modeling systems indicate all human cultural systems, symbol-based modeling processes grounded in language and speech.

Sebeok's tripartite distinction is fundamental in order to distinguish between modeling and communication, as well as to demonstrate the foundational character of modeling with respect to communication.

Language as a modeling device related iconically to the universe it models. This statement was on the same line of thinking as Peirce and Jakobson who both stressed the importance of iconic signs. An equally important connection can be made with Ludwig Wittgenstein's *Tractatus*, particularly with the notion of 'picturing'.

The mind as a sign system or model representing what is commonly called the surrounding world or *Umwelt*. The model is an icon, a kind of diagram, where the most pertinent relations are of a spatial and temporal order. These relations are not fixed once and for all but can be fixed, modified and fixed again in correspondence (a resemblance relation) with the *Innenwelt* (in-

ner world) of the human organism. On the basis of this model which may be compared to a diagram or a map, the human mind shifts from one node to another in the sign network, choosing each time the interpretive route considered most suitable

2. Dialogism

In light of Sebeok's biosemiotics, the concept of dialogism may be extended beyond the sphere of anthroposemiosis and applied to all communication processes. But all communication processes are based not only on *modeling* (Sebeok), but also on *dialogism* (Bakhtin. We believe that modeling and dialogism are at the basis of all communication processes.

Already in *semiosis of information or signification* (Thure von Uexküll), where an inanimate environment acts as a 'quasi-emitter' — or, in our terminology, where the *interpreted* – becomes a *sign* only because it receives an interpretation (an *interpretant*) by an interpreter which is necessarily a leaving being — interpretation is dialogical. Dialogue also subsists in *semiosis of symptomatization* (Thure von Uexküll), where the interpreted too is an interpretant response (*symptom*) that, differentely from *semiosis of information or signification*, does not arise for the sake of being interpreted as a sign. Dialogue subsists also, evidently, in *semiosis of communication* (Thure von Uexküll) where the interpreted itself, before being interpreted as a sign by the interpretant is already an interpretant response calling for interpretation as a sign.

Jakob von Uexküll's 'functional cycle' is a model for semiosic processes. As such, it has a dialogic structure and involves inferences of the 'if...then' type which may even occur on a primitive level, as in Pavlovian semiosis or as prefigurements of the type of semiosis (where we have a 'quasi-mind' interpreter) taking place during cognitive inference.

In the 'functional cycle,' the interpretandum produced by the 'objective connecting structure' becomes an interpretatum and (represented in the organism by a signaling disposition) it is translated by the interpretant into a behavioural disposition which triggers a behaviour into the 'connecting structure.' Uexküll does not use a dialogic model. However , the point is that in the 'functional cycle' thus described, a dialogic relation is established be-

tween interpreted (interpretandum) and interpretant (interpreted by another interpretant, and so forth). The interpretant does not limit itself to identifying the interpreted, but rather establishes an interactive relationship with it.

Vice versa, not only does the 'functional cycle' have a dialogic structure, but dialogue in communication understood in a strict sense, may also be analyzed in the light of the 'functional cycle.' In other words, the dialogic communicative relationship between a sender who intends to communicate something about an object and a receiver may be considered, in turn, on the basis of the 'functional cycle' model. The type of dialogue discussed here corresponds to the processes described by the 'functional cycle' as presented, in Thure von Uexküll's terminology, neither in *semiosis of information or signification*, nor in *semiosis of symptomatization*, but rather in *semiosis of communication*. In this case, the interpreted itself is already an interpretant response before it is interpreted as a sign by the interpretant, this interpreted is addressed to somebody to be identified and to receive the required *interpretant of answering comprehension*.

The *Handbook of Semiotics* by Winfried Nöth (1990) lacks the entry 'Dialogue'. However, this term is listed in the 'Index of subjects and terms,' which informs us that this subject is treated in a chapter titled 'Communication and semiosis' (Part 3). Here the 'functional cycle' is also mentioned (see Nöth 1990: 176-180). This shows the implications of Uexküll's biosemiosic 'functional cycle' concerning the relation between dialogue and communication. Different communication models are discussed in order to show how biological models (such as the models proposed by Maturana, Varela, and Thure von Uexküll) (according to which communication is self-referential, autopoietic and a semiotically closed system) have a dialogic structure and are radically opposed both to the linear (Shannon and Weaver) and the circular (Saussure) paradigms. As reported by Nöth (1990: 180), Thure von Uexküll (1981: 14) demonstrates that Jakob von Uexküll's biosemiosic functional cycle (1982: 8) is characterized by autonomous closure and therefore reacts to its environment only according to its internal needs.

The theory of autopoietic systems is incompatible with dialogism only if dialogism is understood reductively in terms of a communication model

describing communication as a linear causal process. This is a process moving from source to destination. Similarly, there is incompatibility between autopoietic systems and dialogism, when dialogue is conceived as based on the conversation model governed by the turning around together rule. Moreover, the autopoietic system calls for a new notion of creativity. Another question remains, the question of how the principle of autonomous closure is compatible with dialogue conceived as the inner structure of the individual, therefore with creativity and learning.

As Maturana (1978: 54-55) suggests, another form of dialogic exchange may be conceived. This is different from communication conceived as a linear process from source to destination, or as a circular process in which participants take turns in playing the part of sender and receiver. From this point of view, the dialogic model conceived by Maturana may be described in terms of 'pre- or anticommunicative interaction.'

In light of the Bakhtinian notion of dialogism, dialogue is neither the communication of messages, nor an initiative taken by self. On the contrary, self is always in dialogue with the other, that is to say, with the world and with others, whether it knows it or not; self is always in dialogue with the word of the other. Identity is dialogic. Dialogism is at the very heart of the self. The self, 'the semiotic self' (see Sebeok, Petrilli, Ponzio 2001), is dialogic in the sense of a species-specifically modeled involvement with the world and with others. The self is implied dialogically in otherness, just as the 'grotesque body' (Bakhtin 1965) is implied in the body of other living beings. In fact, in a Bakhtinian perspective *dialogue* and *intercorporeity* are closely interconnected: there cannot be dialogue among disembodied minds, nor can dialogism be understood separately from the biosemiotic conception of sign.

As we have already observed (see Ponzio 2003), we believe that Bakhtin's main interpreters such as Holquist, Todorov, Krysinsky, and Wellek have all fundamentally misunderstood Bakhtin and his concept of dialogue. This is confirmed by their interpretation of Bakhtinian dialogue as being similar to dialogue in the terms theorized by such authors as Plato, Buber, Mukarovsky.

According to Bakhtin dialogue is the embodied, intercorporeal, expression of the involvement of one's body (which is only illusorily an individual, separate, and autonomous body) with the body of the other. The image that most adequately expresses this idea is that of the 'grotesque body' (see Bakhtin 1965) in popular culture, in vulgar language of the public place, and above all in the masks of carnival. This is the body *in its vital and indissoluble interconnectedness with the world and the body of others*. With the shift in focus from identity (whether individual, as in the case of consciousness or self, or collective, as in is to say, a community, historical language, or a cultural system at large) to alterity, a sort of *Copernican revolution* as been accomplished. Bakhtinian critique conducted in terms of dialogic reason not only interrogates the general orientation of Western philosophy, but also the dominant cultural tendencies that engender it.

The 'Copernican revolution' operated by Bakhtin in relation to the conception of self, identity, and consciousness involves all living beings and not only mankind. Consciousness implies a dialogic relation that includes a witness and a judge. This dialogic relation is not only present in the human world but also in the biological. Says Bakhtin:

> When consciousness appeared in the world (in existence) and, perhaps, when biological life appeared (perhaps not only animals, but trees and grass also witness and judge), the world (existence) changed radically. A stone is still stony and the sun still sunny, but the event of existence as a whole (unfinalized) becomes completely different because a new and major character in this event appears for the first time on the scene of earthly existence — the witness and the judge. And the sun, while remaining physically the same, has changed because it has begun to be cognized by the witness and the judge. It has stopped simply being and has started being in itself and for itself ... as Well as for the other, because it has been reflected in the consciousness of the other ... . ('From notes made in 1970-71,' in Bakhtin 1986: 137)

At this point, a possible connection may be pointed out between Sebeok's biosemiotic conception and Bakhtin's dialogic conception. These two authors seem very distant from each other. In reality, this is not true. Bakhtin himself was seriously interested in biology. And, in fact, he devel-

oped his own conception of dialogue in close relation to the biological studies of his time, and particularly to the totalizing perspective delineated by Vernadsky and his conception of biosphere. For both Sebeok and Bakhtin, all living beings on the planet Earth are closely interrelated and interdependent, whether directly or indirectly, in spite of their apparent autonomy and separation.

Bakhtinian dialogue is not the result of an attitude that the subject decides to take towards the other. On the contrary, dialogue is the expression of the living being's condition of the biosemiosic impossibility of closure and indifference towards its environment, with which it constitutes a whole system named *architectonics* by Bakhtin. In human beings, architectonics becomes an 'architectonics of answerability,' semiotic consciousness of 'being-in-the-world-without-alibis.' Architectonics thus described may be limited to a small sphere — that is to say, the restricted life environment of a single individual, one's family, professional work, ethnic, religious group, culture, contemporaneity. Or, on the contrary, as consciousness of a 'global semiotic' order (Sebeok), which may be extended to the whole world in a planetary or solar or even cosmic dimension (as auspicated by Victoria Welby). Bakhtin distinguishes between 'small experience' and 'great experience.' The former is narrow-minded experience. Instead

> … in the great experience, the world does not coincide with itself (it is not what it is), it is not closed and finalized. In it there is memory which flows and fades away into the human depths of matter and of boundless life, experience of worlds and atoms. And for such memory the history of the single individual begins long before its cognitive acts (its cognizable 'Self'). ('Notes of 1950,' in Bakhtin 1996: 99)

It must not be forgotten that in 1926 Bakhtin authored an article entitled 'Contemporary vitalism,' in which he discusses problems of the biological and philosophical orders. This article was signed by the biologist Ivan Ivanovich Kanaev, and is an important tessera for the reconstruction of Bakhtin's thought since his early studies. Similarly to the biologist Jakob von Uexküll, Bakhtin too begins with an interest in biology, specifically in relation to the study of signs.

This article by Bakhtin on vitalism was written during a period of frenzied activity, the years 1924-29, in Petersburg, then Leningrad. In this productive period of his life, Bakhtin actually published four books on different subjects (Freud, Russian Formalism, philosophy of language, Dostoevsky's novel). He only signed the last with his own name while the others (together with several articles) were signed by Voloshinov or Medvedev.

In Petersburg Bakhtin lived in Kanaev's apartment for several years. Kanaev contributed to Bakhtin's interest in biology as well as to the influence exerted by the physiologist Ukhtomsky on his conception of the 'chronotope' in the novel. Jakob von Uexküll is also quoted in Bakhtin's text on vitalism.

In 'Contemporary Vitalism,' Bakhtin criticizes vitalism, that is to say, the conception that theorizes a special extramaterial force in living beings as the basis of life processes. In particular, his critique is directed against the biologist Hans Driesch who interpreted homeostasis in the organism in terms of total autonomy from its surrounding environment. On the contrary, in his own description of the interaction between organism and environment, Bakhtin opposes the dualism of life force and physical-chemical processes and maintains that the organism forms a monistic unit with the surrounding world. The relation between body and world is a dialogic one relation where the body responds to its environment modelling is own world.

The category of the 'carnivalesque' – as formulated by Bakhtin and the role he assigns to it in his study on Rabelais – can be adequately understood only in the light of his global (his 'great experience') and biosemiotic view of the complex and intricate life of signs.

The title of Bakhtin's book on Rabelais, literally *The Work of François Rabelais and Popular Culture of the Middle Ages and Renaissance*, stresses the intricate connection between Rabelais's work, on the one hand, and the view of the world as elaborated by popular culture (its ideology, its *Weltanschauung*) in its evolution from Ancient Greek and Roman civilization into the Middle Ages and Renaissance, on the other, which in Western Europe is followed by the significant transition into bourgeois society and its ideology.

Bourgeois ideology conceives bodies as separate and reciprocally indifferent entities. Thus understood, bodies only have two things in common: firstly, they are all evaluated according to the same criterion, that is to say, their capacity for work; secondly, they are all interested in the circulation of goods, including work, to the end of satisfying the needs of the individual. Such ideology continued into Stalinist Russia, which coincides with the time of Bakhtin's writing, and into the whole period of real socialism where work and the productive capacity were the sole factors taken into serious considered as community factors. In other words, work and productivity were the only elements considered as what links individuals to each other. Therefore, beyond this minimal common denominator, individual bodies were considered as being reciprocally indifferent to each other and separate.

The carnivalesque participates in the 'great experience' which offers a global view of the complex and intricate life of bodies and signs. The Bakhtinian conception emphasizes the inevitability of vital bodily contact, showing how the life of each one of us is implicated in the life of every other. Therefore, in what may be described as a 'religious' (from Latin *religo*) perspective of the existent, this conception underlines the bond interconnecting all living beings with each other.

Furthermore, the condition of excess is emphasized, of bodily excess with respect to a specific function, and of sign excess with respect to a specific meaning: signs and bodies — bodies as signs of life – are ends in themselves. On the contrary, the minor and more recent ideological tradition is vitiated by reductive binarism, which sets the individual against the social, the biological against the cultural, the spirit against the body, physical-chemical forces against life forces, the comic against the serious, death against life, high against low, the official against the non-official, public against private, work against art, work against non official festivity. Through Rabelais, Bakhtin recovered the major tradition and criticized the minor and more recent conception of the individual body and life inherent in capitalism as well as in real socialism and its metamorphoses. Dostoevsky's polyphonic novel was in line with the major tradition in *Weltanschauung*, as demonstrated by Bakhtin in the second edition (1963) of his book of 1929.

The self cannot exist without memory; and structural to both the individual memory and social memory is otherness. In fact, the kind of memory we are alluding to is the memory of the immediate biosemiotic 'great experience' (in space and time) of indissoluble relations with others lived by the human body. These relations are represented in ancient forms of culture as well as in carnivalized arts: however, the sense of the 'great experience' is anaesthetized in the 'small,' narrow-minded, reductive experience of our time.

Let us resume. Modeling and dialogism are pivotal concepts in the study of semiosis. Communication is only one kind of semiosis that (together with the semiosis of information or signification and the semiosis of symptomatization) presupposes the semiosis of modeling and dialogism. This emerges clearly if in accordance with Peirce and his reformulation of the classic notion of *substitution* in terms of *interpretation*, we consider the sign firstly as an interpretant, that is to say, as a dialogic response foreseen by a specific type of modeling. Moreover, Bakthin's concept of dialogue also contributes considerably to extending this concept beyond the human world connecting dialogism with semiosis from Sebeok's biosemiotic perspective, namely according to the point of view of *global* semiotics

3. Abduction

Generally one maintains that dialogue commences with signaling behavior from a sender intending to communicate something about an object. But what is not taken into account is that the 'if...then' inference, hypothesis formation, and a 'chain of thought' are dialogic forms in themselves .

In inference, in the hypothetical argument, and in the chain of interpreted and interpretant thought signs generally, dialogue is implied in the relation itself between the interpreted and the interpretant . In inference, in hypothetical argument, and in the chain of thought generally which consists of the relation between interpreted and interpretant signs, dialogue is implied in the relation itself between interpreteds and interpretants. We shall now analyze dialogism in the 'if...then' inference, in hypothesis formation, and in any 'chain of thought'.

The connection between semiosis and interpretation implies the connection between sign and argument, and therefore the connection between semiotics and logic.

Taking Peirce's viewpoint into consideration, we may say that the problem of the connection between identity and alterity in the sign is not just a problem of semiotics, but also concerns logic as theory of argumentation. In Peirce this problem directly concerns logic which as theory of argumentation also involves the problem of dialogue. Considered from the point of view of its relation to the object, the sign is a symbol insofar as it involves mediation of an interpretant; from the perspective of its relation to the interpretant, the sign-symbol is an *Argument*. This is true if the sign-symbol distinctly represents the interpretant which it determines as its Conclusion through a proposition that forms its Premise or, more generally, its premises (see *CP* 2.95). Depending on the type of sign relation that comes to be established in the argument between premise and conclusion, three kinds of arguments are possible: Deduction, Induction and Abduction.

In *deduction* where the relation between the premises and the conclusion is *indexical*, the degree of dialogism is minimal: here, once the premises are accepted the conclusion is obliged.

In *induction* which is also characterized by an unilinear inferential process, the conclusion is determined by habit and is of the *symbolic* order: identity and repetition dominate though the relation between premises and conclusion is no longer obliged.

In *abduction* the relation between premises and conclusion is *iconic* and dialogic in a substantial sense. In other words, it is characterized by high degrees of dialogism and inventiveness as well as by a high margin of risk for error. Claiming that abductive argumentative procedures are risky means that they are mostly tentative and hypothetical with just a minimal margin for convention (symbolicity) and mechanical necessity (indexicality). Therefore, abductive inferential processes engender sign processes at the highest levels of otherness and dialogism. These are the theses we intend to demonstrate in this paper.

The special relation that exists between sign (interpreted) and interpretant as understood by Peirce is a dialogic relation. Peirce evidenced the dialogic nature of the sign and semiosis.

In semiosis of information or signification (cf. Th. von Uexküll, 'Biosemiosis', 1997, and 'Varieties of Semiosis' 1991) where an inanimate environment acts as 'quasi-emitter' — or, in our terminology, where the interpreted becomes a sign only because it receives an interpretation by the interpretant which is a response — receiver interpretation is dialogic. Not only does dialogue subsist in semiosis of communication (Th. von Uexküll) where the interpreted is already itself an interpretant response directed to being interpreted as a sign before it is interpreted as a sign by the interpretant, but dialogue also subsists in semiosis of symptomatization (Th. von Uexküll) where the interpreted is an interpretant response (symptom) which is not directed to being interpreted as a sign, as well as in semiosis of information or signification.

Dialogue does not commence with signaling behavior from a sender intending to communicate something about an object. The whole semiosic process is dialogic. 'Dialogic' may be understood as dia-logic. The logic of semiosis as a whole and consequently of Krampen's semiosic matrix (in the his article entitled 'Model of semiosis' in Posner, Robering, and Sebeok, eds. 1997-2003, I: 248). is dia-logic. The interpretant as such is 'a disposition to respond', an expression used by Krampen (in Posner, Robering, and Sebeok, eds. 1997-2003, I: 259) to describe the dialogic interaction between a sender and receiver.

Krampen's semiosic matrix confirms the connection established between dialogue and semiosis insofar as it shows that the two terms coincide, not only in the sense that dialogue is semiosis but also in the sense that semiosis is dialogue — the latter being an aspect which would seem to escape Krampen. The dialogue process presented in the semiosic matrix is similar to the 'if ... then' semiosic process, to hypothesis formation, chain of thought, and functional cycle after Jakob von Uexküll. In Krampen's article the semiosic matrix illustrates dialogue with two squares which represent two partners, the sender and receiver, where each has its own rhombus repre-

senting the interpretant. Despite this division, the graphic representation of dialogue is not different from the author's diagrams representing other types of semiosis. It could be the model, for example, of an 'if...then' semiosis in which the two distinct interpretants are the premises and the conclusion of an argument in a single chain of thought.

In deduction the relation to the interpretant is of the indexical type; in induction it is symbolic; in abduction, iconic.

In inference the relation between premises and conclusion may be considered in terms of *dialogue* between *interpreted sign* (premises) and *interpretant sign* (conclusion*)*.

The two speakers among whom an argument is hypothetically divided are connected , in deduction, by a relation of reciprocal dependence and constraint. In the deductive argument the premiss determines the conclusion, that is, the precedent determines the consequent with the same force of compulsion with which the past imposes itself upon the present. The conclusion must passively acknowledge the premiss which has already been formulated like a fait accompli (see CP 2.96).

In induction, on the other hand, the conclusion is not imposed by the premiss and is susceptible to modification. Despite this, however, the inductive argument is merely repetitive and quantitative, given that its sphere of validity remains that of the fact, that is, of the totality of facts on whose basis alone can it infer the future. As in deduction, the inductive process is unilinear and moves in a precise order of succession from the point of departure to the point of arrival without interruption, reversal or retroaction as opposed to abduction which, as we will see, moves backwards from the con sequent to the antecedent.

In abduction, the thought-sign (the minor premise) and the thought-interpretant are connected by a dialogic relation which is not pre-determined by the pre-dialogic selection of a law. Retroaction of the interpretant on the premiss to the point that interpretation determines the major premise is precisely what causes us to define this type of reasoning as retroduction or abduction.

The dialogic division between the parts enables us to take into account the level of dialogic complexity, that is, of alterity, differentiation, distance and novelty established in the argument between the sign and the interpretant. The dialogic character of logic is discussed in a medieval tractatus on logic entitled *Summule Logicales* by Pietro Ispano (an author known to Peirce). It is not incidental that Peirce should have used the term *Speculative Rhetoric* to designate *transuasional logic* (see Peirce CP 2.93), the doctrine of the general conditions whereby symbols and other signs refer to and determine the interpretants. In fact, the term Rhetoric implies reference to the addressee, the interlocutor, and recalls such terms as to converse, to argument, to convince and to account for. Furthermore, it represents a break in the conception of reason and reasoning originated from Descartes, and therefore it alludes to the uncertain, probabalistic, and approximative nature of human knowledge. Peircean logic is presented as dia-logic.

The iconicity of abduction consists in establishing a relation between that which originally and naturally is not related: imaginative re presentation attempts an approach to that which is given as other in order to lead it back to a relation of similarity. Similarity is rightly listed by Peirce together with all that we associate with the category of obsistence; in fact, originality or firstness is surpassed by secondness or obsistence when whatever exists autonomously is related to something else. To have an understanding of alterity in a certain sense means to exceed it. The innovating, creative, displacing capacity of abduction is not to be found there fore, so much in its exhibiting an image which draws that which seems to evade all constraints nearer, as in its directing itself towards the autonomously other.

In the abductive process we run the risk of surpassing the datum, thus developing an interpretant that has its own alterity and autonomy in so far as it is not motivated, justified or compensated by the object-datum it specifically refers to. Such self-sufficiency of the abductive interpretant, that is, its iconicity and originality presents a challenge, a provocation with regards to the concept of identity and totality. It thus questions even that which seemed settled and definitive, and exhibits an image which can neither be incorporated nor accounted for whether through immediate reference to the fact

or datum, or on the basis of a system of pre established laws. With a logic that goes beyond the logic of exchange and equilibrium, it is possible for an argument to actualize firstness, originality, or alterity in the very core of the symbolic, of the law, of the transuasional.

In the succession deduction–induction–abduction the *degree of alterity* increases. Says Peirce:

> Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis. Deduction proves that something *must* be; induction shows that something *actually is* operative; abduction merely suggests that something *may be* (*CP* 5.172).

Abduction is the inferential process by which the rule that explains the fact is hypothesized through a relation of similarity (iconic relation) to that fact. The rule acting as general premise may be taken from a field of discourse that is close to or distant from that to which the fact belongs, or it may be invented *ex novo*. If the conclusion is confirmed, it retroacts on the rule and convalidates it (ab- or retro-duction). Such retroactive procedure makes abductive inference risky, exposing it to the possibility of error; at the same time, if the hypothesis is correct, abduction is innovative, inventive and sometimes even surprising.

4. Critique of dialogic reason

If we do not take into account Bakhtins's global (see his notion of 'big experience') and biosemiotic view towards the complex and intricate life of signs, we will not understand the role in his work of the relation between 'dialogism' and 'carnivalesque'. The latter is formulated in Bakhtin's study on Rabelais and then also used in the revised edition (1963) of his book on Dostoevsky.

In *Rabelais* Bakhtin tells us what carnival means for him. He refers to that complex phenomenon, existing in all cultures, formed by the system of attitudes, conceptions and verbal and nonverbal signs according to a comic and joyful idea of living. Carnival, therefore, does not only concern Western

culture, nor the Russian spirit, but any culture of the world insofar as it is human.

Rabelais occupies a place of central importance in Bakhtin's overall conception. By contrast with oversimplifying and suffocating interpretations of Marxism, Bakhtin instead develops Marx's idea that the human being is fully realized when 'the reign of necessity ends'. Consequently, a social system that is effectively alternative to capitalism is one that considers *free time* and not labour time as the measure of *real* social wealth (see Marx, *Grundrisse*, 1857-61, Eng trans.: 708), in Bakhtin's terminology the "time of non official festivity", which is closely connected to what he calls the "great time" of literature.

Today, we are witness to the worldwide spread through global communication of the ideology of production and efficiency. This is in complete contrast with a carnivalesque vision of life. The difference also concerns individualism, which is exasperated by the ideology of production connected with the logic of competitivity. But even when the logic of production, individualism and efficiency is dominant, it cannot eliminate the constitutive inclination of the grotesque body, insofar as it is grounded in dialogism and intercorporeity, for involvement with the world and the body of others. Mankind's inclination for the 'carnivalesque' endures, and this is testified, for example, by literary writing. In Orwell's *1984*, the greatest resistance against a social system based on the ideology of production and efficiency is in fact represented by literature. In this sense we may say that literature (and art, in general) is, and always will be, carnivalized.

Bakhtin's fundamental contribution to semiotics and 'philosophy of language' is his *critique of dialogic reason*, a critiqueconnected to Kant and Marx. Bakhtin inaugurates a 'critique of dialogic reason' by contrast with Kant's 'critique of pure reason' and Sartre's 'critique of dialectic reason'.

Bakhtin privileges the term 'metalinguistics' for his own approach to the study of sign, utterance, text, discourse genre, and relations between literary writing and nonverbal expressions in popular culture, such as the signs of carnival. Bakhtin's critique of dialogic reason focuses on the concept of *responsibility without alibis,* a nonconventional idea of responsibility, which

concerns existential "architectonics" in its relation with the I, the world and others.

Bakhtinian critique of dialogic reason is a critique of the concept of autonomy among individual bodies: in fact, autonomy is an illusion. Consequently, Bakhtin's critique is a critique of individual identity (such as consciousness or self) and of collective identity (such as community, historical language, or cultural system) where identity is conceived in terms of separation from the other following dominant ideological tendencies.

The problem of the critique of dialogic reason leads to the problem of the centrality of dialogue and dialogism in logic, in argumentation and in biosemiosic universe

Bakhtin (in his notes of 1970-71) evidences how unilaterality, ossification, rectilinear and unilateral dialectics derives from sclerotized dialogue. Monologic, unilinear and totalizing dialectics is necessarily orientated towards a synthesis and a conclusion. As such it calls for a "critique of dialogic reason". From this point of view Bakhtin is a milestone because all his research, including his latest paper of 1974 on the methodology of human sciences, focuses on the same problem faced by Sartre in *Critique de la Raison Dialectique*: that is, whether human knowledge and understanding not only imply a specific methods but also a *New Reason*. However, this problem cannot be adequately understood appealing to Sartre's belief in terms of a new relationship between *thought* and its *object*. In fact, Sartre's dialectics remains wholly inside the limits of monologic dialectics for he reduces the relation of otherness to a relation of identity and of reciprocal objectification: dialectics between *for self* and *for others* is dialectics in a totalizing consciousnesses, where the tendency is to assert one's own objectifying view.

> Critique of dialogic reason is critique of the category of Identity which is dominant in Western thought and praxis. From the perspective of identity, sense coincides with partial and limited interests and engenders mystification: and this happens whether we are speaking of the identity of individual, group, nation, language, cultural system or of a macro-community such as the European Community, the Western world, and so fort.

The category of Identity dominates today's world because of the *concrete abstractions* constructed upon it forming the *Reality* we experience: these concrete abstractions are "internal" to today's overall system of social reproduction. They include Individual, Society, State, Nation, Truth, Knowledge, Work, Trade Equality, Justice, Freedom, limited Responsibility, Need, Equal exchange, etc. However, it is not only a question of concrete abstractions ensuing from the system. Even more radically the system itself is grounded in the category of Identity which is asserted structurally and constitutively as the Universal in the worldwide and global processes of Production, Exchange on the Market and Consumption. The logic orienting concrete abstractions in today's processes of social reproduction is the logic of Identity. And the categories of Individual and its rights, obligations, responsibilities, of Society and its interests, of State and its Politics (which reflect Reality as closely as possible), of Equal exchange and its demands, all obey the logic of Identity.

The places of argumentation internal to the order of discourse are the places of the logic of identity. Reason includes 'the reason of war' even if in the form of *extrema ratio,* which presents war as legitimate, just and legal. Reason includes elimination of the other — from emargination and segregation to extermination. Reason is the Reason of Identity. Its logic is asserted by barricading, isolating, expelling or exterminating the other thereby laying the conditions for the construction of the concrete abstractions mentioned above. As anticipated, these concrete abstractions include the category of Individual which must firstly sacrifice its otherness to self in order to assert self as identity.

The Critique of Reason and Argumentation thus understood requires a *point of view that is other*. This approach calls for preliminary *recognition of the other*, or, better, recognition of the fact that recognition of the other is an *inevitable imposition*. Recognition of the other here is not conceived as a concession, a free choice made by the Individual, the Subject, the Same, but as a necessity imposed by alienation, the loss of sense, by the situation of *homo homini lupus*. The situation of *homo homini lupus is consequent* and not mythically antecedent (the allusion is to Hobbes's fallacy!) to such concrete abstractions

as State, Politics, Law. A global semiotic perspective that keeps account of today's socio-economic context in terms of global communication evidences that the human individual, as a living body, is interconnected with all other forms of life over the whole planet thanks to the condition of diachronic and synchronic intercorporeity.

A global and detotalizing approach in semiotics demands availability towards the other, to an extreme degree, a disposition to respond, to listen to others in their otherness, a capacity of opening to the other, where such opening is measured in quantitative terms (the omnicomprehensive character of global semiotics), as well as in qualitative terms. All semiotic interpretations by the semiotician (especially at a metasemiotic level) cannot leave the dialogic relationship with the other out of consideration. Dialogism is, in fact, a fundamental condition for a semiotic approach in semiotics which though oriented globally, privileges the tendency to open to the particular and the local rather than to englobe and enclose. Accordingly this approach privileges the tendency towards detotalization rather than totalization.

As shown by Emmanuel Levinas, otherness obliges the totality to reorganize itself always anew in a process related to what he calls "infinity", and which (to use a phrase associated with Peirce) we could also relate to the concept of "infinite semiosis". This relationship to infinity is far more than cognitive: beyond the established order, beyond the symbolic order, beyond our conventions and habits, it tells of a relationship of involvement and responsibility with the other. This relationship with infinity is a relationship with what is most refractory to the totality, therefore it implies a relationship to the otherness of others, of the other person, not in the sense of another self like ourselves, another *alter ego*, an I belonging to the *same community*, but of an other in its extraneousness, strangeness, diversity, difference towards which we cannot be indifferent despite all the efforts and guarantees offered by the identity of the I.

Such considerations orient semiotics according to a plan that does not belong to any particular ideology. This kind of semiotics concerns human behaviour as it ensues from the awareness of human being's radical responsibility as a "semiotic animal". Properly understood, the "semiotic animal"

is a responsible actor capable of *signs of signs*, of mediation, reflection, and awareness in relation to semiosis over the whole planet. In this sense global semiotics must be adequately founded in cognitive semiotics, but it must also be open to an other dimension that is the ethical. This is the pivotal object of what Susan Petrilli and myself propose the to call '*semioethics*'. Its commitment is the "health of semiosis" that is *of life* on Planet; and to cultivate an understanding of the entire semiosic universe, semiotics, as *semioethics*, must continuously refine its auditory and critical capacity, that is, its capacity for listening and criticism.

Translation from Italian by Susan Petrilli

References

Bakhtin, Mikhail M., 1929. *Problemy tvorchestva Dostoevskogo* (Problems of Dostoevsky' work). Leningrad: Priboj. It. trans. and Intro. M. De Michiel, *Problemi dell'opera di Dostoevskij*. Intro. A. Ponzio. Bari: Edizioni dal Sud, 1997.

—, 1963. *Problemy poetiki Dostoevskogo*, Moscow, Sovetskij pisatel'. Eng. trans. and ed. C. Emerson, *Problems of Dostoevsky's Poetics*. Minneapolis: University of Minnesota Press, 1984.

—, 1965. *Tvorchestvo Fransua Rable* Moscow: Khudozhestvennia literatura. Eng. trans. H. Iswolsky, ed. K. Pomorska, *Rabelais and His World*. Cambridge: The Massachusetts Institute of Technology, 1968.

—, 1981. *The Dialogic Imagination: Four Essays*, ed. C. Emerson and M. Holquist. Austin: University of Texas Press.

—, 1986. *Speech Genres & Other Late Essay*, ed. C. Emerson and M. Holquist. Austin: University of Texas Press.

—, 1990. *Art and Answerability*, ed. M. Holquist and V. Liapunov. Austin: University of Texas Press.

—, 1993 [1920-24]. *Toward a Philosophy of the Act*, Eng. trans. V. Liapunov, ed. M. Holquist. Austin: University of Texas Press.

—, 1996. *Sobranie sochinenij* [Collected works], Vol. 5 (1940-1960), ed. S. G. Bocharov and L. A. Gogotishvili. Moskow: Russkie Slovari.

Bakhtin, Mikhail M.; Valentin N. Voloshinov, 1929. *Marksizm i filosofija jazyca*. Leningrad. Eng. trans. L. Matejka and I. R. Titunik, *Marxism and the Philosophy of Language*. Cambridge (Mass.): Harvard University Press, 1973. It. trans. *Marxismo e filosofia del linguaggio*, ed. A. Ponzio. Lecce: Manni, 1999.

Bouissac, Paul, ed., 1998. *Encyclopedia of Semiotics*. New York: Oxford University Press.

Bouissac, Paul; Michael Herzfeld; and Roland Posner, 1986. *Iconicity. Essays on Nature and Culture. Festschift for Thomas A. Sebeok on the his 65th birthday*. Tübingen: Stauffenburg.

Clark, Katerina; Michael Holquist, 1984. *Mikhail Bakhtin*. Cambridge: Harvard University Press.

Hardwick, Charles S., ed. (in collab. with J. Cook), 1977. *Semiotic and Significs. The Correspondence Between Charles S. Peirce and Victoria Lady Welby*, Intro., ix-xxxiv. Bloomington-London: Indiana University Press.

Heidegger, Martin, 1927. *Being and Time*, Eng. trans. J. Macquarric and E. Robinson. New York and Evanston: Harper & Row, 1962.

Holquist, Michael, 1990. *Dialogism. Bakhtin and his World*. London and New York: Routledge.

Krysinski, Wladimir,1984. Bakhtine et la question de l'ideologie. *Etudes Françaises* 20: 21- 36.

—, 1998. *La novela en sus modernidades. A favor y en contra de Bajtin*. Madrid: Iberoamericana.

Levinas, Emmanuel ,1961. *Totalité et Infini*. The Hague: Nijhoff. Eng. trans. A. Lingis, *Totality and Infinity*, Intro. J. Wild. Dordrecht-Boston-London: Kluwer Academic Publishers, 1991.

—, 1974. *Autrement qu'être ou au-dela de l'essence*. The Hague: Nijhoff. Eng. trans. A. Lingis, *Otherwise than Being or Beyond Essence*. Pittsburgh: Duquesne University Press, 2000.

—, 1991. *Entre nous. Essais sur le penser-à-l'autre*. Paris: Grasset. Eng. trans. *Entre nous. On Thinking-of-the-Other*. London: The Athlone Press, 1998.

Lotman, Jurij M., 1984. O Semiosfere. *Trudy po znakovym sistemam* 17, 5-23.

Lucid, Daniel P., ed. , 1977. *Soviet Semiotics. An Anthology*. Baltimore: John Hopkins University Press.

Maturana, Humberto N., 1978. Biology of Language: The Epistemological Reality. In *Psychology and Biology of Language and Thought*, ed. G. A. Miller and E. Lenneberg, 27-63. New York: Academic Press.

—, 1980. Autopoiesis: Reproduction, Heredity, and Evolution. In *Autopoiesis, Dissipative Structures, and Spontaneous Social Orders*, ed. M. Zeleny, 45-107. Boulder: Westview Press.

Maturana, Humberto R.; Francisco J. Varela, 1980. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel.

Morris, Charles, 1971. *Writings on the General Theory of Signs*, ed. T. A. Sebeok. The Hague-Paris: Mouton.

Nöth, Winfried, 1990. *Handbook of Semiotics*. Bloomington: Indiana University Press.

Orwell George, 1949. *Nineteen Eigthy-Four. A Novel*. New York: Penguin, 1982.

Peirce, Charles S., 1923. *Chance, Love and Logic*, ed. Morris R. Cohen. New York: Harcourt.

—, 1931-1966. *Collected Papers of Charles Sanders Peirce*, eds. C. Hartshorne, P. Weiss, and A. W. Burks, 8 Vols. Cambridge (Mass.): The Belknap Press, Harvard University Press. (References are to *CP*, followed by volume and paragraph numbers).

Petrilli, Susan, 1998a. *Su Victoria Welby. Significs e filosofia del linguaggio*. Naples: Edizioni Scientifiche Italiane.

—, 1998b. *Teoria dei segni e del linguaggio*. Bari: Graphis. New ed. 2001.

—, ed. 2003a. *Translation Translation*. Amsterdam . Rodopi.

—, ed. 2003b. *Linguaggi*, Bari: Laterza.

— ed. 2004. *Ideology, Logic, and Dialogue in Semioethic Perspective*,

Semiotica. Journal of the International Association for Semiotc Studies, special

issue, 148-1/4.

– (2005). *Percorsi della semiotica*. Bari: Graphis.

Petrilli, Susan; and Patrizia Calefato (2003). *Logica, dialogica, ideological. I segni fra funzionalità ed eccedenza*, Intro. S. Petrilli, 9-14. Milan: Mimesis.

Ponzio, Augusto, 1970. *Linguaggio e relazioni sociali*. Bari: Adriatica, new ed. Bari: Graphis, 2005.

—, 1986. *Dialogo sui dialoghi* (in coll. with Massimo A. Bonfantini). Ravenna: Longo.

—, 1990. *Man as Sign*. Eng. trans by S. Petrilli. Berlin: Mouton de Gruyter.

—, 1993. *Signs, Dialogue and Ideology*. Engl. trans. S. Petrilli, Amsterdam: Benjamins.

—, 1998a. *Signs of Research on Signs* (in collab. with S. Petrilli): Special Issue of *Semiotische Berichte* , 22 / 3, 4, 1998,

—, 1998b. *La revolución bajtiniana. El pensamiento de Bajtín y la ideología contemporánea* (1998),

—, 2000a. *Philosphy of Language, Art and Answerability in Mikhai Bakhtin* (in collab. with Susan Petrilli). Ottawa : Legas.

—, 2000b, Three dialogues on rhetoric, argumentation, and media (in collab with Massimo A. Bonfantini and Susan Petrilli), *Semiotica*, 128-1/2, 69-112.

—, 2001. Thomas Sebeok and the Sign of Life (in collab. with S. Petrilli). Cambrige: Icon Books.

—, 2002. *I segni e la vita. La semiotica globale di Thomas A. Sebeok* (in collab. with Susan Petrilli). Milan: Spirali.

—, 2003a . *Semioetica* (in collab. with Susan Petrilli), Rome: Meltemi.

—, 2003b. *Tra semiotica e letteratura. Introduzione a Michail Bachtin*, Milan: Bompiani

—, 2005a. *Reasoning with Emmanuel Lévinas* (in collab. with Susan Petrilli and Julia Ponzio). Ottawa: Legas.

—, 2005b. *Semiotics Unbounded. Interpretive Routes in the Open Network of Signs* (in collab. with Susan Petrilli), Toronto: Toronto University Press,.

—, 2005c. *Semiotic Animal* (in collab. with Susan Petrilli and John Deely). Toronto: Legas.

Posner, Roland.; Klaus Robering; and Thomas A. Sebeok, eds. (1997-2004). *Semiotik Semiotics. A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, 4 Vols. Berlin: Walter de Gruyter, 1998.

Rossi-Landi, Ferruccio, 1973. *Ideologies of Linguistic Relativity*. The Hague: Mouton.

—, 1975a). *Linguistics and Economics*. The Hague: Mouton. New ed., 1977.

—, 1975b. Signs about a Master of Signs. *Semiotica* XIII, 2, 115-197. Now in F. Rossi- Landi 1992a, 17-57.

—, 1975c [1953]. *Charles Morris. Una semiotica novecentesca.* Milan: Feltrinelli.

—, 2005 [1978]. *Ideologia.* Milan: Mondadori. Eng. trans R. Griffin, *Marxism and Ideology*. Oxford: Clarendon, 1990.

—, 1992. *Between Signs and Non-Signs*, ed. Susan Petrilli, Amsterdam: J. Benjamins.

Rudy, Stephen, 1986. Semiotics in the Ursss. In The Semiotic Sphere, ed by T. A. Sebeok and J. Umiker-Sebeok, 34-67, New York: Plenum.

Sartre, Jean-Paul, 1960. *Critique de la raison dialectique*. Paris: Gallimard.

—, 1961 [1943]. *L'Etre et le néant*. Paris: Gallimard. Eng. trans. H. E. Barnes, *Being and Nothingness*, Intro. M. Warnock. London: Routledge, 2001.

Sebeok, Thomas, A., 1991. *A Sign Is Just a Sign*. Bloomington-Indianapolis: Indiana University Press. It. trans. and Intro. S. Petrilli, *A Sign is just a sign. La semiotica globale.* Milan: Spirali, 1998.

—, 1994. *Signs. An Introduction to Semiotics*. Toronto: Toronto University Press. New ed. 2001.

—, 2001. *Global Semiotics*. Bloomington: Indiana University Press.

Sebeok, Thomas A.; and Susan Petrilli (1998). Women in Semiotics. In G. F. Carr *et alii* eds., *Interdigitations: Essays for Irmengard Rauch*. Bristol, England: Thoemmes Publisher.

Sebeok, Thomas A.; Susan Petrilli; and Augusto Ponzio, 2001. *Semiotica dell'io*. Rome: Meltemi.

Uexküll, Jakob von, 1909. *Umwelt und Innenwelt der Tiere*. Berlin: Springer Verlag.

—, 1940. *Bedeutungslehehre*. Leipzig: Verlag von J. A. Barth.

Uexküll, Thure von, 1981. The sign theory of Jacob von Uexküll. In Classic of Semiotics, ed. by Martin Krampen, et al: 147-170. New York: Plenum.

Uexküll, Thure von; and Wolfgang Wesiack, 1988. *Theorie der Humanmedizin: Grundlagen ärtzlichen Denkens und Handelns*. Munich: Urban & Schwarzenberg..ernadsky, Wladimir I., 1926. *Biosfera*. Leningrad: Nauka.

Welby, Victoria, 1983 [1903]). *What is Meaning? Studies in the Development of Significance*. Ed. and Pref. A. Eschbach, ix-xxxii, Intro. G. Mannoury, xxxiv-xlii. In *Foundations of Semiotics,* Vol. 2. Amsterdam-Philadelphia: John Benjamins.

—, 1985 [1911*Significs and Languag*, ed. and Intro. H. W. Schmitz, ix-ccxxx-vii. In *Foundations of Semiotics,* Vol. 5. Amsterdam-Philadelphia: John Benjamins.

—, 1985b. *Significato, metafora, interpretazione*, It. trans., ed. and Intro., 7-50. S. Petrilli. Bari: Adriatica.

Wellek, René,1991. Mikhail Bakhtin. In *A History of Modern Criticism*, VII. New Haven:Yale University Press.

Abduction and metaphor
Tropism of truth

José Augusto Mourão
Departamento de Ciências da Comunicação
Universidade Nova de Lisboa

For the methods of thinking that are living activities in men are not objects of reflexive consciousness".
Charles S. Peirce, C.P. 3. 404, 1892

Abductive inferential processes engender sign processes at the highest levels of otherness and dialogism.
S. Petrilli/A. Ponzio

Metaphor, dream, parable, allegory, the whole of art, the whole of science, the whole of religion, the whole of poetry, totemism, the organization of facts in a compared anatomy, – they are all examples, our clusters of examples, of abduction, inside the sphere of the human mind.
G. Bateson

Abduction and Metaphor

The scientific culture does not progress just by logical inferences, but also by lateral extensions of knowledge due to parabolic and metaphorical phenomena (Mary Hess). That means that both inference and metaphor contribute to increase knowledge. According to René Thom 'all metaphors are true'. It is the aim of this paper to connect metaphor, as a means to prove that the system works, with abduction, always working along the same

lines as Gregory Bateson who ascribes to this frame of reference a unique connotation.

Apparently, only the poets take metaphor seriously. Musil does it, in his paper *Rede zur Rilke* of January 1927, taking the metaphorical (*das Metaphoryche*), in its highest degree.[1] As we all know, in the metaphor, the properties of one thing are used to refer to something else, that is, this something else is defined by what it's not, as happens in the "quiet light" of the *Duino Elegies*, by the late Rilke. The question of Massimo Cacciari: "is the metaphor a sign for unsettled restlessness?" satisfies a great deal of what constitutes the metaphorical process, this being, eventually, the making of a new state of restlessness in language.[2] Umberto Eco said some twenty years ago that "Literature and nature do not tell the same story (1981:18)". The "as if" situation, that defines literature, gives metaphor the sufficient distance to dematerialize what's material. To R. Barthes the logic of "effect of realness" (Barthes 1994: 479) points the deconstructionist critics to any reference to the immediate reality. Literature isn't reality but it is through it that we make inferences regarding what we call "reality", by analogy with the state of affairs and mental states contained in the world of the text. The question is in acknowledging what's real in what's fiction. "The actual task is not the identification of reality as a symbolic fiction, but to show that there is something in that symbolic fiction that is more than fiction."[3]. Fiction is but a figure taken for real and treated as an effective event. To figure (from *fingere*) signifies to form, to invent, to fake. A fiction is an "amplified form of a figure", "a figure is a small fiction", "a figure is a sketch of fiction" and "the metaphor, in a more general way, the figure or, at least, the figures by substitution, like metaphor or metonymy, antiphrasis, litotes or hyperbole, are verbal fictions and miniature fictions"[4].

1 R. Musil, Essays und reden, in Gesammelte Werke, vol. VIII, pp. 1229 ss.

2 Massimo Cacciari, *paraíso y naufragio. Musil y el Hombre sin Atributos*, Madrid, Abada, 2005, p. 80.

3 Slavoj Zizek, *Arriesgar lo imposible. Conversaciones con Glyn Daly*, Madrid, Trotta, 2006, p. 99.

4 Gérard Genette, *Metalepsis*, Reverso, Barcelona, 2005, p. 16, 17.

Over the metaphors hangs the accusation that most of them are false or that, in essence, the metaphor is nothing but an *ars combinatoria*, lacking evidence value: "the sentences in which metaphors occur are true or false in a normal, literal way, for if the words in them don't have special meanings, sentences don't have special truth (Davidson, 1984: 257)". Metaphor would be no more than a rhetorical mechanism or a linguistic ornament. Nevertheless, it was forgotten that both metaphor and metonymy are cognitive instruments. Metaphors and metonymies, generalized, conventionalized, and lexicalized (improperly said to be dead) are, from the cognitive point of view, the most important (Silva, A. Soares da 1997: 73). The classical studies of Lakoff & Johnson (1980) show that we systematically conceptualize many domains of experience through conceptual metaphors, that is, projecting in them other domains. For example, if we conceptualize a religion through the metaphor RELIGION IS WAR, we get two domains (religion and war) analogically bonded, so that the followers of one religion are like the opponents in war. Another famous example: we think and speak of human life as a voyage, more exactly as three voyages (cf. Lakoff & Turner 1989): birth, the travel throughout the world (we are travellers and we must make decisions), and death (we leave, we departure). In other words, the "metaphor is not merely a semantic extension (or transference) from one category to another of a different domain, but it involves a systematic and coherent analogy between the internal structure of two domains of experience and, therefore, all relevant knowledge associated to the concepts and domains in question." (Silva: 75). Imagetic schemes (Johnson 1987 e1 Lakof 1987, 1990) and perspectivation (Taylor 1995: 90, 125.6) are concepts that allow us to situate processes, actions, attitudes, emotions.

The literal meaning is the relation we establish between a word and an object, or, better said, between a concept and a sound-image. What's particular in metaphor is the change in this relation, in a way that, in a given moment, a concept can have a literal meaning and a metaphorical one. Here lies the reason why the metaphor is a remarkable phenomenon, creator of meanings. The learning we have acquired of the word *virtue* and its derivatives in statements like:

"Philip has many virtues"

leads us to say that the statement is true if, and only if, Philip has intellectual or manual virtues (or gifts) that are acknowledged (he plays the piano well) or even moral virtues (he is a man of his word). The statement is true if Philip has "good operative habits". We shall say that the statement is false if we don't believe these qualities are present in him or, if in fact, he has none of these qualities. For a Latin of the classical period, the statement would be false if among Philip's virtues there weren't virility and such characteristics which the Latin speaker believes to be fit of a man, like the value (Chamizo Domínguez, 1999). In another context, the meaning of *virtus* can be different, thus changing its value of truth. The acceptance of the truth contained in a basic adequation – metaphorical or literal – between a term and an object allows us to make inferences and to build nets of concepts that shall be considered as true or false by the speakers of a given language, according to its conformity with the adequation where they stand. Here, coherence works as a basic presupposition – analogous with the principle of charity – causing us to suppose that any discourse is rational and reasonable.

The principle underlying metaphor is not only rhetorical or linguistic, it is a semiotic one. The extension of knowledge also occurs through the metaphor and both scientists and epistemologists know this. Not only have the experiments in laboratories and explicit inferential systems produced extensions of knowledge. There are mental experimentations that can also achieve this.

There are small narrative structures that have a value as cognoscitive experiences[5]. Another question: are there pronominal metaphors? Fabbri uses the narrative structure of the sower. Through this parable, Jesus proposes a metaphor: the act of listening. But the same metaphor can also be used to explain what language is, what to a linguist would be a lousy metaphor[6].

The several readings that arise from this parable are what Bateson calls *abduction*. In *Mind and nature,* this author says that abduction is not only an inferential movement, but mainly a deeper but lateral understanding of what

5 Paolo Fabbri, *El Giro Semiótico*, Barcelona, gedisa, 2000, p. 92.

6 *Op. cit.*, p. 89.

a metaphor or a parable is. Abduction would consist in the case of the seeds, in asking the metaphor to think, to know how to use the figurative in order to form a chain of inferences.

From analogy to abduction

In all great thinkers there seems to be a "before" and an "after, a first phase (as young) and second (as matured). So it is the case of Charles Peirce. But it is not easy to make a complete and rigorous reconstruction of the concept of abduction in Peirce. The author, in his "Collected Papers" makes several references to this concept, describing the implication, in all levels and in all phases of his theoretical approach, mainly when he attempts to determine the fundamental principles of his philosophy (the feneroscopic categories), when he conceives the logical and semiotic elementary mechanisms and, finally, when he ends his investigation towards pragmatism. In 1878, he published in Popular *Science Monthly* the paper "deduction, induction, hypothesis", which represents the classic source where the researchers of his work find the three types of inference[7]. To show the contrast between several kinds of thinking, Peirce uses the example of beans, not always very well understood. (CP 2.621-623, 1878). To characterize the inference of a case from a general rule and a result, he calls it, not abduction, but hypothesis or conjecture. As in the case of induction, hypothetical inference doesn't have a necessary feature, only probable, being also a synthetic or amplified thinking.

Peirce provides examples from the natural sciences (from the presence of sea fossils, one can infer that there was ocean in that site previously) and the human sciences (from the documents that make reference to Napoleon, one can well infer that Napoleon never existed). In the ideas of the matured Peirce, abduction is a kind of inference characterized by its probability: the conclusion achieved is always a conjectural one, probable, that might seem plausible to the researcher. It is in this intuitive feature that lays its validity and not in its effective probability which has only an indirect influence. (CP

7 Tradução castellana de Juan Martín Ruiz-Werner em C. S. Peirce, "Deducción, inducción e hipótesis", Aguilar, Buenos Aires, 1970.

2.102, 1903). Here, the author speaks about "retroduction" and "abduction". It is worth to quote one of the texts where the concept of abduction is more evident:

> "Abduction is that class of operation that suggests a statement that is not, in any way, contained in the data preceding it. There is another name to it; it is abduction, because it is no more no less *guessing*. A given object presents an extraordinary combination of characters to which we would like to give an explanation. The number of names in the directory is not even close to the multitude of possible laws of attractions which could be have been explained by Kepler's laws of motion of planets and which, before verification by the prediction of disturbances, etc, would have explained them perfectly.
>
> Newton, one might say, supposed that the law would be one and simple. Nevertheless, how did he come to this, if not by accumulating guesses upon guesses? Surely, there are far more complex phenomena than simple ones. (...)"

The strength and weakness of the abductive thinking is located on the level of the relation settled between the two following premises: one that states a result, usually observation data, and another that states a rule, a principle, a general condition, or a cause that might provide an explanation to this result, and, by doing so, transforming it in a case of general application. The relation between result and rule is, as one can deduct, a necessarily hypothetical and conjectural relation. The role of abduction is to provide an explanatory hypothesis. It will be necessary for these consequences to be objectively verifiable. The fecundity of abduction is connected to the power of prevision brought by deduction; abduction and deduction frame induction. But it is induction that corroborates the previous two. Without hypothesis and deduction, induction would be reduced to an empty form of knowledge

Abduction and science

All knowledge proceeds by inference; all knowledge begins with hypothesis and ends with observations framed by hypothesis: these observations simulate the readjustments of hypothesis, therefore new observations, and so on *ad infinitum*. To Popper, the scientific method has two faces: an inven-

tive and a demonstrative. If the first helps to formulate good hypothesis, the second is mostly about proving hypothesis (Popper 1973). Abduction in Peirce is mostly a bet on the abstract power of the intellect. For that reason, he recurs to Aristotle and the scholastics. It is through this abstraction process that we capture the laws and the universality, what's accessory and what's essential. This ability is an intuition of the universal laws of things of nature and of society. (CP 5.173). Mary B. Hesse has already stated that the scientific culture doesn't only progress through logical inferences, but also through lateral extensions of knowledge, due to parabolic or metaphorical phenomena.

To Peirce, there is not a single discovery in the scientific theory – whether the exact sciences, whether the human sciences – that is not due to the abductive process. Abduction can be understood as an inclination to believe that a given hypothesis is plausible or probable, with some risks of it becoming wrong, but also with a few chances of becoming true in a given episteme (which includes both knowledge and belief) in a given historical time. Peirce aligns it as the counterpoint in man of what is the instinct for the animal, a kind of plausibility of some hypothesis. The "abductive instinct" would be as the lowest realm of abduction. It is not a thought but a feeling, from a basic presupposition: "nihil est in intellectus quod non prius fuerit in sensu".

What do we talk about, when we talk about scientific discovery? In normal speech, we use the term discovery in such broad situations as these:

1. Mary has discovered her glasses inside the fridge.
2. Opening the safe, Philip discovered a treasure.
3. Ana discovered the shop that sells the glasses she wants.
4. Teresa discovered that God exists.
5. Christopher Columbus discovered America.
6. Mendel discovered the laws of inheritance.

In each of these statements, the word "discovery" has a different meaning and usage. Etymologically, to discover means to remove what's covering or hiding[8]. Such sense is only found in the first two phrases, although in

8 *Etymological Dictionary* by Adolfo Coelho, Port. Edition.

different senses: the whereabouts of an object is discovered but nothing actually new is added. The opening of a safe, though, reveals the existence of an acquired concept: that of a treasure, although it would be necessary to consider the contents of the safe in order to identify that concept. In the third statement, Ana has an *a priori* idea: the object that is looked for must conform to the idea of the glasses she is looking for. It is not a precise object but, rather, an object among others, with which it shares some properties. If the discovery of God can be shared with others, the discovery of America was made, as history says, by one single individual: Columbus. Only the last statement could be a "scientific discovery". Mendel chose the pea plant as the model organism and points out reasons, of biological and practical nature, for his choice. He formulates hypotheses to explain the results he obtains and invents ways to test them. He states principles to which the phenomenon of inheritance must obey, in conformity with the results of his qualitative and quantitative analysis. Mendel conceives, actually, a duality: genotype vs fenotype, proposing a non-linear correspondence between them. Mendel had to invent them in order to explain the phenomena of life evidenced by the manipulation which he previously planned. Mendel doesn't remove the veil that covers the laws of inheritance; rather, he uses a living being, incomprehensible. The analysis corresponds to a dissection. Mendel constitutes a remarkable case of scientific method and thought.

It is, therefore, the role of abduction to introduce new ideas in science. The beginning of a research is always abductive. It's the hypothesis that leads the way, showing which experiments should be carried out and the direction in which we should be looking. The classical authors said the beginning of all research was wonder. Wonder in face of the astonishing phenomena is the beginning of knowledge, said Aristotle.[9] Aristotle relates analogy and hypothesis in the same process, starting from similarity (*Top.* I, XViii, 108b). Metaphor by analogy is one of the four different kinds (*Poet.* XXI, 7-14, 1457b), being the best and the most popular. It is not possible to make ad-

9 Aristóteles, *Metafísica*, I, 982b 12-17.

equate abductions without using analogy. This is to Peirce a forth kind of inference:

> "Analogy is the inference that a collection not very great of objects that share a number of aspects might, very likely, share another aspect. For example, the Earth and Mars, share so many aspects that it doens't seem unlikely that they share the fact that they are both inhabited." (CP 1.69, 1896).

To discover and to detect are, according to Peirce, two complementary ways of obtaining an increase of information. Discussing the differences between discovery and detection, he defends that "discovery is the finding out of a new thing – as the discovery of America – while the finding out of a new character should be called a detection" (CP 2.430). But, in fact, detection and discovery have something in common: they are both different kinds of abductive inference, "the only kind of reasoning which supplies new ideas, the only kind of which is, in this sense, synthetic" (CP 2.777). So, the "logic of detection" (as logic of discovery") is based on an abductive inference.

Analogy is crucial to generalize, inside a margin of vagueness, "to which a kind of analogous universal is required, not purely univocal and not purely equivocal", says M. Beuchot[10]. Analogy in the shape of copies and metaphors is the mode to invent the same in the other (CP 2.277, 1903), although abduction and analogy are different, as M. Beuchot has noted. Analogy is between similarity and difference, the difference being the predominant factor (*Ibidem*: 63). Analogy is thus bounded to the iconical sign that "shows a similarity or analogy with the theme of the discourse" (CP 1.369, 1885). Analogy would be a combination of inductive inference and hypothetical inference. The operativeness of analogy lies, according to Johansen, in its capacity to connect different semantic and cognitive domains and to use that connection in the production of knowledge, whether trivial or innovative[11].

10 Mauricio Beuchot, *art. cit.* p. 67.

11 Jörgen Dines Johansen, "La analogía y la fábula en literatura", in Analogía filosófica, XII, 1998, nº 1, p. 78.

Science and Literature

In reaction to the scientific triumph, post-modern thought tends to present science whether as a mere structure of power or as another kind of literature. Deconstruction of science for the sake of literature, itself going through deconstruction, doesn't help to define the proper functioning of these two social and human practices. The obvious connection between scientific abduction and literary creativity (Anderson 1987) doesn't allow, however, reducing science to literature. Carontini talks of a very important difference between the role of abduction in scientific discovery and its role in the textual interpretation, more particularly in a novel. "The textual genre is usually animated by a unified narrative project. In most cases, the narrative universe is an orderly and closed universe. It is, therefore, possible to observe, by trial and error, an hypothesis that can account for other interpretations. On the contrary, the knowledge of a man of science, just as our common daily knowledge, cannot rely on the rigorous and logical organization of a text. The universe of our knowledge is not a closed universe; it doesn't obey to a plot.

To defend us from the "infinite causality of the real" we only dispose of abduction which, in this case, remains an inference eminently fallible"[12]. In other words, what distinguishes the textual interpretation from other forms of knowledge is that, in the first case, the recipient can expect a more or less coherent discourse. He knows, moreover, that the author of the text makes use of an encyclopaedia that is more or less common to both. That's part of the writing protocol. But this is not the case in other kinds of knowledge. In the case of the scientific discovery, the researcher presupposes that "the knowledge universe with which he is confronted, obeys to a project comparable to that of the textual universe. Unlike the reader of a text, the degree of creativity of the activity of the knowledge subject can be radically innovative and create explanatory hypotheses that are truly different and new, in relation to his own universe of knowledge." (*Ibidem*: 225). The inven-

12 Enrico Carontini, "Le rôle de l'abduction dans le processus d'interprétation » in L. Sfez e G. Coutlé, *Téchnologies et symboliques de la communication* PUG 1990, p. 224-225.

tion that implies scientific knowledge is, in a way, present in the metaphor, which is a kind of abduction. "Every authentic metaphor presupposes that the interpreter can operate transferences can be able to shift an explanatory principle from one semantic field to another. Metaphor is a risky operation, eminently fallible and profoundly conjectural but that don't block communication." (*Ibidem*: 226). In fact, textual interpretation implies a true creative work. But the explanatory principle with which science works is something radically new in relation to the universe that is known to the subject. Here, to recover a metaphor of Peirce, we need a "lume naturale", and not just a "lumen culturale". In this point, Bateson and Peirce coincide. Bateson writes: "What seems to happen is that in nature, there are, in a corresponding manner, reflected in our processes of thought, vast regions where the abductive systems prevail. For example, the anatomy and physiology of the human body can be considered as a vast abductive system, with its own coherence... Similarly, the environment, inside which the creature lives, constitutes an abductive system internally coherent, although this system is not immediately coherent in relation to the organism"[13].

Metaphor / Abductive Principle

Let's reflect on how the metaphor communicates with the abductive process. Abduction is the type of inferential process through which an hypothesis is built. One can well say, as Peirce did in 1901, that "I make an abduction every time I express in a phrase what I see" (MS 692). Here, the function that explains the fact is turned into hypothesis based on the relation of similarity (an iconical relation) with the fact. If the conclusion is confirmed, it acts retrospectively over the function and validates it. (hence, ab-duction or retro-duction). It is this kind of retroactive process that makes abductive inference risky, exposing it to the possibility of error. When the hypothesis is correct, abduction is innovative and surprising. (cf. Bonfantino 1987). Peirce wrote: "Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely in-

13 Bateson, op, cit., p. 17.

volves the necessary consequences of a pure hypothesis. Deduction proves that something *must* be; induction shows that something *actually* is operative; abduction merely suggests that something *may be'* (CP 5.172). The close ties between abductive inference and verisimilitude come from the fact that, as Victoria Welby has shown, 'one of the most splendid of all our intellectual instruments' is the 'image or the figure' (1985ª) (1911: 13; cf. also Petrilli 1995ª, 1998ª).

The dialogic nature of the signs in inference and in hypothetical argument has been studied, particularly by Ponzio (1990: 197-214). The abductive process reaches high levels of dialogism "and generates responses of the riskiest and most inventive and creative kind" (S. Petrilli/A. Ponzio, 2005: 26). That is, its proceeding mode is always by trial and error, hypothetical and provides only the minimal for (symbolic) convention and the mechanical necessity (indexicality). The dialogic relation between who interprets and that which is interpreted is determined by the connection between abduction and iconicity. In deduction, the degree of dialogism is minimal: here the relation between premises and conclusion is of an indexal order: when the premises are accepted, the conclusion is compulsory. In induction, we are also faced with an linear inferential process, in which the conclusion is determined by habit and is of a symbolic order: here identity and repetition prevail, even if the relation between premises and conclusion is not compulsory. In abduction, on the contrary, the relation between premises and conclusion is dialogical and iconical in a substantial sense. In the terms by Ponzio: "It is characterized by high degrees of dialogism and inventiveness as well as by a high risk margin of error" (Ponzio, 2004: p. 51). As is known, in abduction, the relation between premises and conclusion is inconic – that is, it is a degree of reciprocal autonomy, which can lead either to a high degree of inventiveness or a margin of risk of error. To Peirce, a sign conveys to another by inferences; Greimas does it through metaphor or narrative. P. Fabbri, one of Greimas's most acute disciples, tries to convert these two cognitive strategies by removing the two biggest obstacles that arise from the rigid distinction between proof and syllogism on one hand and, from the conception of

metaphor as a mere ornament, on the other. One of the greatest virtues of *La svolta semiotica*, is the way Fabbri relates metaphor with cognition, in terms of enunciative phenomena. Here, the metaphor is taken not only in terms of statement or sign (a woman is a rose) but in terms of enunciation. The most interesting is to know what happens when we say "I am my Mondays". What relation is there between "I" and "Monday"?

Surprising Facts

Abduction communicates with the logic of discovery, the process through which one generates hypotheses to give account of those facts that surprise us. Abduction as a process of finding explanatory hypotheses sets out a "surprising phenomenon". As Peirce wrote: "Our knowledge of any subject *never goes beyond collecting observations and forming some half-conscious expectations, until we find ourselves confronted with some experience contrary to those expectations... [A]t once we turn over our recollections of observed facts; we endeavour so to rearrange them, to view them in such new perspective that the unexpected experience shall no longer appear surprising. (CP 7.36).*

The effort to explain a surprising fact supposes that the facts which are observed are part of a more vast system of facts. (CP 7.36). This "larger system" is not, however, more than "a conjecture or a guess" (CP 7.36). The surprise produces a certain irritation and demands an hypothesis, an abduction, which makes the surprising phenomenon normal and reasonable. What is, then, the logic structure of abduction? We follow the 7th of his lessons on pragmatism. (CP 5.189, 1903):

> There is a surprising fact C;
> However, if A was true, C would be a *matter of course*.
> Therefore, there are reasons to suspect that A is true.

The surprise is the detonator of any research, the shock with an anomaly. Jaime Nubiola, studied the 127 occurrences of the word "surprise" in Peirce's *Collected Papers* and mainly in texts after 1901[14]. "Each branch of

14 Jaime Nubiola, "La abducción o lógica de la sorpresa", in Numero Actual, febrero-Abril 2001, p. 8.

science begins with a new phenomenon that violates some kind of negative subconscious expectation" (CP 7.188, 1901). However, what makes a phenomenon surprising is not mere regularity. "Nobody is surprised that the trees in a forest do not make a regular set. Mere regularity doesn't cause our surprise, because irregularity in our lives is normal. (CP 7.189,1901):

> "Mere irregularity, where no definite regularity is expected, creates no surprise nor excites any curiosity. Why should it, when irregularity is the overwhelmingly preponderant rule of experience, and regularity only the strange exception? In what a state of amazement should I pass my life, if I were to wonder why there was no regularity connecting days upon which I receive an even number of letters by mail and nights on which I notice an even number of shooting stars! But who would seek explanations for irregularities like that" (CP 7.189, 1901).

Seen in depth, detective stories are all built in this way. The very medical diagnosis of surprising symptoms and types of diseases that make such symptoms reasonable are excellent examples of effective use of abduction in our daily life[15]. Metaphors can be considered, according to Max Black, as a kind of model: "Metaphor is a submersed model". (Ortony 1993: 30) and, for this reason, the secondary theme of the metaphor offers a perspective through which the primary theme is seen (Achilles is seen in the perspective of the lion).

This is the meeting point between parable and metaphor. The capacity to make conjectures is neither blind nor infallible but it is an instinctive capacity, analogous to the animal instincts that enable a bird or a bee to fly. (CP 6.476m 1908).

From the three classes of elementary thinking that Peirce describes – abduction, deduction and induction – abduction is essential to understand scientific discoveries. (CP 8.209, c. 1905). What distinguishes these three classes of thinking are not the rules but the function of these inferential forms in the scientific processes. Neither abduction nor neither induction have any pretension to be truthful. Michael Hoffmann (1998: 44-45) points out a specific form of abductive inference:

15 U. Eco e T. Sebeok, eds. *El signo de los tres*, Lumen, Barcelona, 1989.

(F1) (1) The surprising fact F is observed.

(2) but if H was true, then F would be a matter of course ,

So,

(3) There are reasons to believe that H is true. (cf. CP 5.189, 1903).

How can we find the hypothesis H? Peirce "identifies abduction as guessing, considering this "guessing" as an instinctive power, on one hand, and a process that operates on the basis of information under our control, on the other." (Kapitan 1991: 8). In Peirce's epistemology all cognition – from perception to the logical and mathematical thinking – is mediated by signs of general elements and by contexts. But, if every cognition is mediated by some general elements, one should supose the same for the "surprising facts", in the sense that they too, are perceptible.

The surprise only exists related to certain expectations and in some circumstances. (cf. CP 2. 776, 1901-02; 6.469, 1908). If, according to Peirce, abductive inference reflects itself in the perceptive judgement, without a clear separation between them (CP 5. 181, 1903), then we can say that every abduction in face of a surprising fact is nothing more than the search for a mode of perception for that fact. (cf. CP 6.469, 1908).

Code

Massimo Cacciari distinguishes the metaphorical from the analogical. Quoting Novalis, he writes: "Analogy is antithetically synthetic", while the metaphor is always separable, an apparent agreement, analogy establishes an inseparable difference and expresses, *in uno,* difference and inseparability.

The metaphor "masks" the difference which, in the end, must be reaffirmed." (pp. 81-82). If analogy is symbolic, metaphor is dia-bolic. Analogy doesn't tend to comparison; it starts with the weighting of the difference between the dimensions of the beings in opposition. It names things by what they are not. (op. cit., p. 81). Taking metaphor seriously should mean two things: either in the sense of its dissolution, or in the sense of its overcoming in the form of an analogy. Moreover, Cacciari says that what moves the

desire for this last one is the desire for the unreal: the expression of what's singular, the form of the whole. "A metaphor compares things without explaining the comparison. It takes what is true about a number of things and applies them in another group of things. When we say that a nation is "decomposing" we use a metaphor that suggests that certain changes in a nation are alike those in a fruit rotting with bacteria. But we do not mention the fruit or the bacteria."…"But the complete list of metaphors or images clings to earth, which wouldn't happen in a dream", writes Gregory Bateson[16]. Wanting to measure the exactness of a metaphor is to destroy it. Metaphor only works when we hide its "infinite evil" of continuous remission, allusiveness, or enthusiasm which grows over it[17]. An analogy without intuition, without that vision, is, in itself, empty. Any research is opposed to the principle of immediate intuition. Johansen finds in the tick-tock of a clock the same formal structure of an analogy:

a : b :: a: b.

Parallelism is, to Jakobson, the most recurrent poetical resource (Jakobson in Sebeok 1960: 358). This principle applies to the phonological, syntactical and semantic levels. Analogy, which is seen as a process of knowledge that works according to the same inference mode of abduction, can contain metonymy (from the fragment to the whole) and metaphor (the tension between the literal and the figurative senses). Furthermore, Peirce himself allows this closeness when he admits the process of abductive knowledge in the creating of analogies between things. (Beuchot, 1998: 57). "Genuine analogy requires negation; if it is not to become a deadening univocity, and dialectical language must move forward into analogy if anything but a chaotic equivocity is to be left when "the purging fire" of negation ends (Tracy, 1981, 417-20). Both are aspects of analogical imagination which tries to formulate a theological comprehension of the original, in relation to the realities of God, of the Self and the world. "Because to the mystic the analogy is also a perfect event, a gift alone." (Cacciari: 87). Unity in separa-

16 Gregory Bateson, *Metadiálogos*, Lisboa, gradiva, sd. p. 94-95.

17 Cacciari, op. cit., p. 83.

tion, such is the realm of the metaphorical. Isn't this the very definition of love?

Bibliografia

Anderson, Douglas R. (1987) *Creativity and the Philosophy of C. S. Peirce*, Dordrecht, Nijhoff.

Bateson, Gregory*, Metadiálogos*, Lisboa, gradiva, sd. p. 94-95.

Eco, Umberto (1981), "Guessing: from Aristote to Sherlock Holmes", *Versus*, nº 30, 1981.

Lakoff, George & Turner, Mark (1980) *Metaphors We Live By*, Chicago, The University of Chicago Press.

Lakoff, George & Turner, Mark (1989) *More than Cool Reason: A Field Guide to Poetic Metaphor*, Chicago, The University of Chicago Press.

Petrilli S./A. Ponzio (2005), *Semiotics Unbounded*, Univ. Of Toronto Press.

Popper, Karl (1973). *La logica de la investigación científica*, Madrid, Tecnos.

Silva, Augusto Soares da (1997) A Linguística Cognitiva, Revista Portuguesa de Humanidades vol. 1-1/2, Faculdade de Filosofia da U.C.P. Braga.

David Tracy (1981) *The analogical imagination*, SCM Press, New York.

Vilela, Mário (1996) "A metáfora na instauração da linguagem: teoria e aplicação", Revista da Faculdade de Letras da Universidade do Porto – Línguas e Literaturas 13, pp. 317-356.Abduction, Fallacies and Rationality in Agent-Based Reasoning

Abduction, Fallacies and Rationality in Agent-Based Reasoning

Lorenzo Magnani
Department of Philosophy and Computational Philosophy Laboratory
University of Pavia
lmagnani@unipv.it

Elia Belli
Department of Philosophy and Computational Philosophy Laboratory
University of Pavia
elia.belli@unipv.it

1. Agent-based reasoning, agent-based logic, abduction

It is well-known that in classical logic a good argument is a sound argument and, from a semantic point of view, it is a valid argument based on true premises. Even if this conception of good inference is usually able to model many kinds of argumentation of real human beings, its appeal to true premises is ill suited to many contexts which are often characterized by the presence of hypothetical and uncertain beliefs, by great disagreement about what is true and false, by ethical and aesthetic claims which are not easily categorized as true or false, and, finally, by variable contexts in which dramatically different assumptions may be accepted and rejected.

We share with Gabbay and Woods (2005) the idea that logic can be considered a formalization of what is done by a cognitive agent. Starting from this perspective, logic is *agent-based* (Gabbay and Woods, 2001). In this perspective logic has to describe cognitive agent's ways of reasoning: if reasoning is an aid to cognition, a logic, when conceived of as a theory of reasoning, must take this cognitive orientation deeply into account.

According to Norman and Hutchins's view (Norman, 1993; Hutchins, 1995), we can say that a *cognitive system* is a triple of a cognitive agent, cognitive resources, and cognitive target performed in real time. Correspondingly, a logic of a cognitive system, or an *agent-based logic*, is a principled description of conditions under which agents deploy resources in order to perform cognitive tasks. The more general notion involved in this conception is that reasoning is target-motivated and resource-dependent. That is, agent-based logic deals with the problems, limitations and ways of reasoning of cognitive agents.

In this perspective the problem on describing how agents perform reasoning is constrained in three crucial ways: in what they are disposed towards doing or have it in mind to do (i.e., their *agendas*); in what they are capable of doing (i.e., their *competence*); and in the means they have for converting competence into performance (i.e., their *resources*). Loosely speaking, agendas are plans of action, for example belief-revision and belief-update, decision-making and various kinds of case-making and criticism transacted by argument (cf. Gabbay and Woods, 2005).

A description of the cognitive practice should include a description of the *type of cognitive agent* involved. Agency-type is set by two complementary factors. One is the degree of command of resources an agent needs to advance or close his (or its) agendas.

Agent–based reasoning consists in describing and analyzing the reasoning occurring in problem solving situations where the agent access to cognitive resources encounters limitations such as

bounded information
lack of time
limited computational capacity

So, for cognitive agendas, three types of resources are especially important and they are (1) *information*, (2) *time*, and (3) *computational capacity*.

Hence, the "beings-like-us" that Woods describes in his "Epistemic bubbles" (Woods, 2005) discharge their cognitive agendas under press of incomplete information, lack of time, and limited computational capacity.

We can consequently say that cognitive performances depend on information, time, and computational capacity. An *agent-based logic*, as a discipline that furnishes ideal descriptions of *agent-based reasoning*, returns to be thought of as a science of reasoning and considered agent–centered, task–oriented, and resource–bound. Woods says:

> So, then, a principal function of reasoning is to facilitate cognition, this means the reasoning agent is also a cognitive agent. If logic is to press forward as a renewed science of reasoning, it would do well to reflect on what cognitive agency is like, on what it is like to be a knower (Woods, 2005, p. 732).

In dealing with these features we arrive to what has been called the "Actually Happens Rule" (Woods, 2005, p. 734) that states that "to see what agents should do we should have to look first at what they actually do and then, if there is particular reason to do so, we would have to repair the account". This rule is a particular attractive assumption about human cognitive behavior mainly for two reasons. The first is that beings like us make a lot of errors; the other is that cognition is something that we are actually very good at.

In the following section we will discuss the case of "fallacies" as errors that people make. These errors occur in ways of reasoning and acting that from some perspectives are good and from others are bad. In dealing with this matter we will try to give an account of fallacies seen from the viewpoint of agent–based reasoning. We will try to give some examples of fallacious reasoning treating both informal fallacies (such as the inductive ones like "hasty generalization") and formal fallacies(such as abduction). We will treat induction(in the case of hasty generalization), and abduction as fallacious ways of reasoning that in spite of their fallacious character are fruitful for the cognitive agent: away of being rational through fallacies.

Abduction can be easily considered in the perspective of agent-based reasoning because in abductive reasoning (Magnani, 2001 b) both the activity of guessing new explanatory hypotheses (or new "reasons"), and the activity of selecting already existing ones, is based on incomplete information. In this case we deal with "nonmonotonic" inferences: we draw defeasible

conclusions from incomplete information.[1] From this perspective, abductive reasoning also represents a prototypical case of practical reasoning: we adopt deliberations based on incomplete information and on particular abduced hypotheses – guesses – that serve as "reasons."

2 Beings-like-us as hasty generalizers

As already noted, people make errors in reasoning. This means that in analyzing the beings–like–us argumentations we have to face problems regarding agent's access to cognitive resources such as information, time, and computational capacity, and logical attributes such as truth–preservation. It is in this sense that we have previously said that agents discharge their cognitive agendas under press of bounded information, lack of time and limited computational capacity.

The successful use of fallacies into many kinds of reasoning can be fruitfully accounted for in the framework of *agent-based reasoning*. It is undeniable that in human reasoning mistakes are widespread. The peculiarity of fallacies seen in the perspective of agent–based reasoning is that mistakes that are actually committed are mistakes that do not seem to be mistakes to those who commit them. In some sense we can say that they are ways of reasoning that are felt truth preserving for the reasoner but are not considered truth preserving for the logicians!

A fallacy is a pattern of poor reasoning which appears to be a pattern of good reasoning (Hansen, 2002). Fallacies are forms of reasoning and argumentation typical of organic agents and in this sense we can say they are suitably shaped by evolution. Simple inductions and abductions performed more or less consciously by both humans and animals are surely two great results of this evolutionary process. Two main disciplines respectively clearly illustrate different kinds of fallacies: formal logic, which recognizes and explains "formal fallacies", and informal logic, that describes the so-called "informal fallacies". First of all, we can say that the validity of a deductive

1 It has to be noted that an explanatory reasoning can be causal, but explanations are also based on other aspects. Thagard (1992, chapter five) illustrates various types of explanations (deductive-nomological –so-called in the neopositivist tradition of philosophy of science – statistical, causal, analogical, schematic).

argument depends on its form, consequently, formal fallacies are arguments which have an invalid form and are not truth preserving (for example the fallacy of the "affirming the consequent" and of "denying the antecedent"). On the other hand, informal fallacies are any other invalid modes of reasoning whose failing is not strictly based on the type of the argument (for example the "*ad hominem* argument" or the "hasty generalization").

Even if there is no agreement upon an established taxonomy, the fallacies discussed in *informal* logic contexts typically still include formal fallacies (of course also discussed in *formal* logic) such as the famous fallacy of affirming the consequent[2] and the fallacy of denying the antecedent, but also the proper informal fallacies such as "*ad hominem*"(against the person), "slippery slope", "*ad baculum*" (appeal to force), "*ad misericordiam*" (appeal to pity), and "two wrongs make a right".

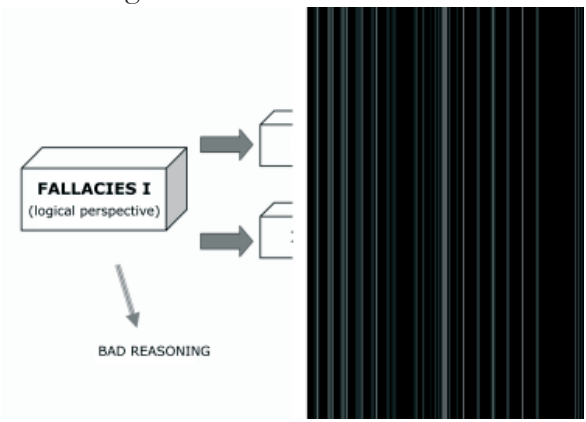


Figure 1. Fallacies from the point of view of classical logic.

From the point of view of classical logic a fallacy is a bad argument that looks good (cf. Figure 1). From the point of view of agent-based reasoning a fallacy is not an argument that looks good but is bad, but an argument that is bad in some aspects and good in some others. We can say that a fallacy is a mistake made by an agent. It is a mistake that seems not to be a mistake. Correspondingly, it is a mistake that is naturally made, commonly made, and not easy to repair (i.e., to avoid repeating) (Woods, 2004, ch. 1).

2 In the light of classical logic abduction coincides with the fallacy of affirming the consequent (cf. below section 3).

Let us consider the inductive case of the so–called "hasty generalization", that can lead the cognitive agent – in spite of its fallacious character – to fruitful outcomes.

This fallacy occurs when a person (but there evidence of it also in animal cognition, for example in mousse) infers a conclusion about a group of cases based on a sample that is not large enough. It has the following form:

- Sample *S*, which is small, is taken from the group of persons *P*
- Conclusion *C* is drawn about the group *P* based on S

It could take also the form of:

- The person X performs the action A and has a result B
- Therefore all the actions *A* will have a result *B*

The fallacy is committed when not enough A's are observed to warrant the conclusion. If enough A's are observed then the reasoning is not fallacious, at least from the *informal* point of view. Males, when driving their cars, have probably quarreled with a woman driving her car and, while quarreling, they have argued (when not shouted) "all woman are bad drivers!" This is our case of fallacious reasoning.

Insofar, small samples will be likely to be unrepresentative. Another simple case is the following. If we are asking one person that even recently met a lot of Italians what he thinks about the recently new established Italian proportional electoral system, his answer clearly would not be based on an adequate sized sample for determining what Italians in general think about the issue. This is because the answer given is based only on a reduced experience and that judgment cannot be relevant in dealing with a generalization about the matter in question. This means that this fallacious argument implies that small samples are less likely to contain numbers proportional to the whole group of cases.

People often commit hasty generalizations because of bias or prejudice. For example, someone who is a sexist might conclude that all women are unfit to fly jet fighters (or to drive a car) because one woman crashed in either case. People also commonly commit hasty generalizations because of laziness or sloppiness. It is very easy to simply jump to a conclusion and much

harder to gather an adequate sample and draw a justified conclusion. Thus, avoiding this fallacy requires minimizing the influence of bias and taking care to select a sample that is large and meaningful enough.

Moreover, we can recognize another important occurrence. We have said that people commit errors and are hasty generalizers because of prejudice, mindlessness, bias, and so on. What we are trying to underline is that the hasty generalization is not always a bad generalization for two reasons. The first is that, getting true conclusions, hasty generalizations might be good if the result of the generalization we made coincides with the *result* of a good generalization in the philosophical sense of induction, or in the sense of inductive logics. We call this case "casual" truth preserving feature of hasty generalization. The second reason is that, in some sense, even if we do not reach good conclusions, not exploiting the casual truth preserving feature, we can say that hasty generalization is good in some sense, obviously not in the classical logic one. We will now try to understand what it can be.

Think of a toddler that for the first time touches a stove in his kitchen (Woods, 2004). His finger is now burned because the stove burns. Starting from this evidence, the hasty generalizer toddler thinks that all the stoves are hot and decides not to touch stoves anymore. This is obviously a hasty generalization:

X of observed *A* are *B* (The stove *touched* burns)

Therefore all *A* are *B* (All the stoves burn)

Or:

Sample *S*, which is too small, is taken from the group of persons *P,* (The toddler touches the stove and at a first touch the stove burns).

Conclusion *C* is drawn about the group *P* based on *S*. (All the time the toddler will touch the stove, it will burn).

We can also say that this is a case of bad argument also from the formal point of view because it is not truth preserving, in the light of classical logic. However, in the perspective of agent-based reasoning the problem now is: can we say that this argument is good from some perspective? Indeed the hasty generalization is sometimes a "prudent" strategy. It also presents a

cognitive economy: given the task of not being burnt for a second time, the hasty generalization is a kind of reasoning that is fruitful because, being a prudent strategy, it embeds the canons of *strategic rationality* in the sense of the "strife for survival". Moreover, it also involves a *cognitive success*.

Hence, it is clear that fallacies (hasty generalization in this case) have some relevant relations with strategic rationality. However, the prudent strategy of "not touching the stove" is obviously incorrect for at least two reasons: 1) the first reason is that it is not good to generalize from only one sample available and 2) the second reason resorts to the fact that the causal relationship between the fact that a stove burns if it is overheated is not discovered by the toddler. From applied natural physics, we can say that it is a state of affairs that a simple stove(not overheated) does not burn because a stove is made of iron or some other metals and metals burn only if they are overheated. So there is something "bad" in this kind of *fallacious* reasoning certainly – first reason – from an informal logic point of view but also – second reason – from the perspective of natural physical principles of heat, that constitute the only epistemologically satisfactory knowledge of the situation. But even if we acknowledge these wrong aspects, there is an idea of some rationality embedded in this example due to the fact that the toddler prevents himself from being burned. It seems that hasty generalization (like in the case of other fallacies, such as the fallacy of affirming the consequent can be considered resources that form a sort of *human survival kit* (Woods, 2004, p. 7). As some unconditioned reflexes, hasty generalization is a response (in the form of a reasoning and then of a possible subsequent action) to something that the toddler is involved to. The cognitive result of a hasty generalization is bad but only in the sense that it is based only on one sample and because it does not *explain* the burnt stove. It is instead a form of good reasoning because it preserves the toddler from being burnt another time.

We have also to note that hasty generalization also allows the toddler to produce a new *successful cognitive information*. In the perspective of the logical tradition, this piece of information is "bad" because obtained through fallacious reasoning, but in agent-based terms we notice that the same infor-

mation contributes to solve the toddler's problem and, in this sense, can be endowed with "good" cognitive relevance.

The goal reached by the toddler through the hasty induction is in some sense a success. In describing agent–based reasoning, we have said that agent–based logic (as a matter analyzing agent-based reasoning) returns to be thought of as science of reasoning and considered agent-centered, task-oriented and resource-bound. The task of the toddler was to prevent himself from being burnt: the task is reached with a bad performance, in the perspective of classical logic. We can say that his fallacious strategy "satisfies" his purpose of not being burnt; we can also say not only that hasty generalization halts upon the attainment of the good–enough, but also reaches a specific target that responds to a survival strategy. In some sense, the toddler's reasoning in hasty generalization terms can be seen as a very good strategy that satisfies his purpose in an optimizing way. From this point of view, the study of fallacies allows us to think about the interplay between reasoning, strategy, goal, and evaluation of goals (cf. Figure 2).

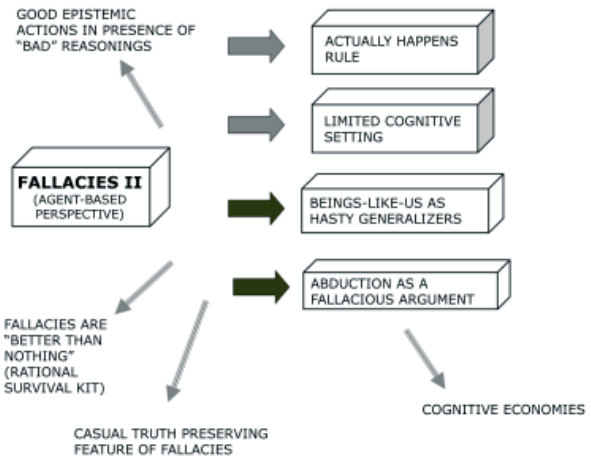


Figure 2. Fallacies from the point of view of agent-based reasoning. The big dark arrows show the examples of fallacious reasoning and the other two arrows underline the setting in which we are moving. The thin arrows display the possible implication of fallacious reasoning considered in agent-based term.

For a better understanding of the matter, we can follow the Kirsh and Maglio's dichotomy: they distinguish actions into two categories: *pragmatic actions* and *epistemic actions* (Kirsh and Maglio, 1994). Pragmatic actions are the actions that an agent performs in order to arrive as close as possible to a goal. The environment in which the agent moves is modified so that it assumes a configuration that helps the agent to reach the physical goal. On the other hand, epistemic actions are the actions that an agent performs in the environment in order to discharge the mind of a cognitive load or to extract information that is hidden or that would be very hard to obtain only by internal computation (cf. Figure 3).

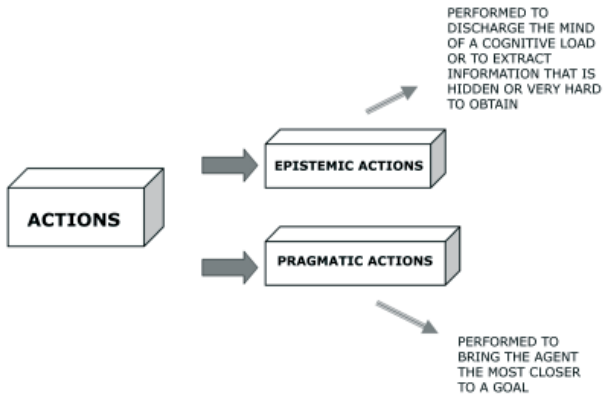


Figure 3. Epistemic and pragmatic action.

The toddler's action of touching the stove is a pragmatic action performed by the toddler in order to reach a goal. For example, we can think that the baby was trying to reach some biscuits on the fridge and in trying to reach them, he touched the burning stove. But, on the other side, the toddler's action is also an unexpected but welcome epistemic action. In fact, it furnishes the basic information on which the inductive hasty generalization can be based. As we have seen in the perspective of the logical tradition the reached generalization is "bad" because it is a form of fallacious reasoning, but in agent-based terms the same information contributes to solve the toddler's problem of not being burnt for a second time and,

in this sense, can be endowed with "good" cognitive relevance. In other words, we can say that, in this case, the action that is the basis of a fallacious reasoning is a good epistemic action, but also a good pragmatic action (in the case the toddler has previously reached the biscuits on the fridge).

We are trying to say that fallacious forms of reasoning can be good not only for practical purposes, such as *task-oriented reasoning*, in solving here and-now problems. The question we are gathering is if action deriving from fallacious reasoning can receive an epistemic endowment. In fact we can say that, from the situation previously experienced through the "touching the stove" action, the toddler obtains new information and creates a strategy that prevent himself of not being burned more seriously for a second time. His fallacious reasoning(hasty generalization, in this case) suggests that he should not touch the stove for at least two reasons. Practically because this touching action has already given him bad results, and theoretically because he made a logical inference that governs his way of future acting, no matter if the inference is bad from a classical logical point of view. We maintain that fallacies also deal with epistemic action, in the sense that they can exploit them to extract new information.

We have contended above that fallacies are forms of reasoning and argumentation typical of organic agents and in this sense we have concluded they are part of a "survival kit" suitably shaped by evolution. We have also added that induction and abduction performed more or less consciously by both humans and animals are surely two great results of this evolutionary process. We know that in the last centuries humans have also characterized induction(and abduction)in various "ideal" philosophical and logical ways, so going beyond the spontaneous use of those kinds of thinking we have just illustrated. Already Mill provided "Methods" for Induction and Peirce integrated abduction and induction through the famous syllogistic framework where the two non–deductive inferences can be clearly distinguished: it has to be noted that Mill also said that what he called "institutions" rather than individuals are the embodiment of "inductive logics". Following this Millian perspective (Mill,1843), Gabbay and Woods also add that it is typical of human individuals to function as practical agents and that it is typical of

"institutions" to function as theoretical agents (Gabbay and Woods, 2005, p. 14); moreover, agents tend toward enhancement of cognitive assets when this enables the achievement of cognitive goals previously unaffordable or unattainable.

In summary, organic agents like human beings are hasty generalizers and more or less naive inducers and abducers but are also the creators of external cognitive and logical representations that for example provide deductive and computational realizations of those reasoning performances. The interplay between these "external" tools and the already "internalized" templates of reasoning certainly realizes a continuous improvement of the internal templates themselves but also expresses the centrality of the hybrid exploitation of both levels in reasoning.

3 Abduction as a fallacious argument

It is well–known that abduction appears to be a formal fallacy that can be recognized from the classical logic point of view: the fallacy of affirming the consequent (cf. below) (Peirce, 1931–1958). However, from the point of view of both everyday and scientific knowledge, abduction is an important kind of inference used to explain facts and invent hypotheses and theories (Magnani, 2001b).

Abduction –a distinct form of reasoning –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, 2001b): 1) abduction that only generates "plausible" hypotheses ("selective" or "creative") and 2) abduction considered as inference "to the best explanation", which also evaluates hypotheses (cf. Figure 4). An illustration from the field of medical knowledge is represented by the discovery of a new disease and the manifestations it causes which 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.

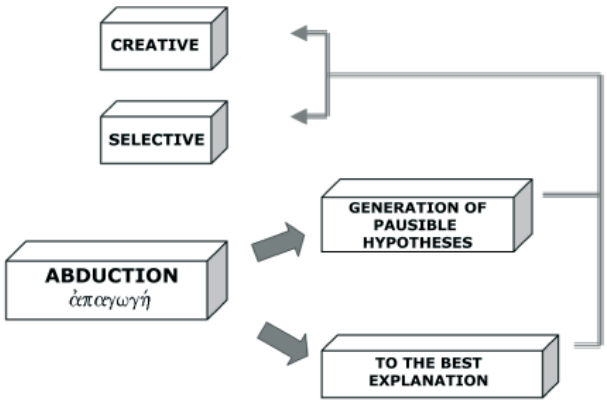


Figure 4. Creative and selective abduction.

Magnani (2001b) has introduced the concept of *theoretical abduction* as a form of neural and basically internal processing. 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. Figure 5).

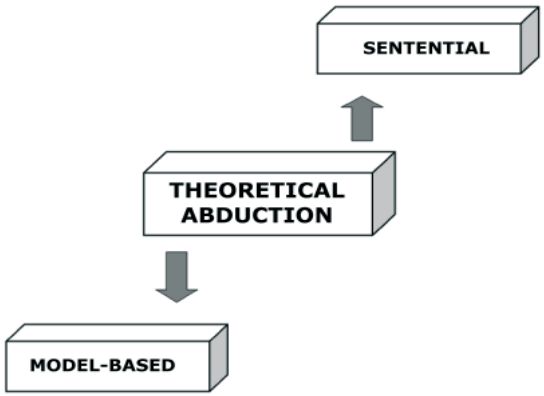


Figure 5. Theoretical abduction.

Theoretical abduction 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 called *epistemic mediators* (Magnani, 2001b). The concept of *manipulative abduction*[3] captures a large part of scientific 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.

Let us now illustrated the fallacious character of abduction. 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: by 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 am 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" (Hempel, 1966) and is related to the idea of confirmation of scientific hypotheses, predominant in neopositivistic philosophy but also present in the anti-inductivist tradition of falsificationism (Popper,1959).

3 Manipulative abduction and epistemic mediators are introduced and illustrated in (Magnani, 2001a) and (Magnani, 2001b).

Deduction is an inference that refers to a logical implication. Deduction may be distinguished from abduction and induction on the grounds that the truth of the conclusion of the inference is guaranteed by the truth of the premises on which it is based only in deduction. 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:

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.

(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 generalization, like the piece of hematologic knowledge represented by (1) (in this case we meet induction as the ability to generate simple laws, contrasted with 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)

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

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

If a patient is affected by a beta-thalassemia, his/her level of hemoglobin A2 is increased.

John is affected by a beta-thalassemia.

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 no abduction because there is no "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

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

It is now clear that in the light of classical logic abduction is a fallacy. Despite this fallacious character it is of fundamental importance in many agent-based reasoning situations like scientific explanation, scientific discovery, and also in moral deliberation, as we will illustrate below. We can furnish another reason that stresses the fruitfulness of abduction in agent-based reasoning: it is a powerful inferential process able to govern inconsistencies. For example, in the case of the formation of new scientific hypotheses and theories epistemologists have always recognized the role played by inconsistencies and anomalies that violate the paradigm–induced expectations derived from previously established conceptual frameworks. Logicians have in turn shown that inconsistencies generated by anomalies are difficult to be

managed in deductive frameworks: they are unexpected facts that the rules of classical logic are not able to manage.

In fact, if we consider *deduction* in scientific reasoning, an anomaly in the system is a problem that logical hypotheses will explain with difficulty: *classical* deduction has no sense in giving reasons for unexpected or unknown facts. In a deductive system, all premises are given and new information that contradicts the previous one clashes against the classical law of *ex falso sequitur quod libet*.

On the contrary, abduction starts from the philosophical assumption that information is incomplete and this retroductive method shows that it is a fallacy if considered in the framework of classical deduction. In other words, in a classical logic system, adding new information to the old one does not let us to revise previous assumed hypotheses. But if we start from incomplete information we will have to face the problem of inferring conclusions that are not foreseeable or already stated and that in the meantime contradict the previous ones. As a consequence of this we obtain that previous deduced conclusions can be invalidated and modified in presence of a new information. For these reasons logicians quickly recognized that abduction is a form of *non-monotonic* reasoning that only non–standard logic environments can grasp. Recently Gabbay and Woods (2005) have nicely reminded us that in the reasoning situations characterized by limited information reasoning cannot be anything but a fruitful *ignorance preserving reasoning*. In a logic able to deal with this fundamental aspect we have to clearly distinguish between the formation of hypotheses and the cogency of their validation.[4]

Stating this, we can outline two different ways of perceiving abduction: 1) from the point of view of classical logic, abduction is a formal fallacy, not truth preserving; 2) from the point of view of everyday reasoning and epistemology, abduction is an important kind of reasoning able to discover new hypotheses and also to furnish explanations of scientific facts. In delineating the structure of a new agent–based perspective of logic able to encompass fundamental human reasoning devices like abduction Gabbay and Woods

4 An example of ignorance–preserving reasoning is diagnostic reasoning in which abduction is useful plausible diagnoses and deduction for its validation.

further point out that logic has to be considered an "account of how thinking agents reason and argue" (2005, p. 1). As we have already mentioned their idea is that logic has to be considered the disciplined description of the behavior of real–life logical agents. Logic has to be thought of as an *agent-based logic*. From this viewpoint, abduction can be rendered as that kind of logical reasoning in which the fact of not being truth preserving (but *ignorance preserving*), has to live together with the fact that it is fruitfully used by real logical agents. Similarly induction is seen as *probability enhancing* and deduction as *truth preserving*.

The description of this new logical perspective is based on the assumption that the reasoner reasons in abounded information set typical of real agents and consequently on the redefinition of the significance of many forms of reasoning previously considered fallacious.

The new agent-based perspective on logic makes the logical system more open and flexible. In this sense the study of fallacies, as important tools for the human "kit" providing evolutionary advantages as well, allows us to maintain that, in some sense, a fallacious form of reasoning is better than nothing (Woods, 2004). In the case of abduction, we can go beyond this definition and say that fallacious reasoning give us even more: in fact, the use of abduction is good for at least two reasons. Abduction is not only a simple formal fallacy, but also a specific case of ignorance–preserving agent's reasoning that can be fruitfully idealized in theoretical logical agents; on the applicative side, abduction is a good process able to provide new hypothesis and govern inconsistencies.

Conclusion

In this paper we have described the relation between logical fallacies, induction, and abduction considered as inference to the best explanation. We have exploited the so–called *agent-based reasoning* framework, which adopts the perspective of a cognitive agent, to show how the so–called strategic rationality is important in many human reasoning performances. We have described how in the perspective of agent-based reasoning *fallacious* reasoning can be redefined and considered as a good way of reasoning. A quick

illustration of the main aspects of abduction has been introduced showing its character from the point of view of classical logic, which depicts it as a kind of formal fallacious argument –the well–known fallacy of affirming the consequent. We have stressed the strategic role of abduction as a method of explanation and discovery in science and in everyday reasoning (for example abduction is able to govern inconsistencies that arise in scientific domains, when anomalies trigger the generation of new hypotheses).

Reference

[Gabbay and Woods, 2001] D. Gabbay and J. Woods. The new logic. *Logic Journal of the IGPL*, 9(2):141–174, 2001.

[Gabbay and Woods, 2005] D. M. Gabbay and J. Woods. *The Reach of Abduction*. North-Holland, Amsterdam, 2005. Volume 2 of *A Practical Logic of Cognitive Systems*.

[Gabbay and Woods, 2006] D. M. Gabbay and J. Woods. A formal model of abduction. In L. Magnani, editor, *Abduction, Practical Reasoning, and Creative Inferences in Science*, volume 14 (2), pages 189–220, 2006. Special Issue of the *Logic Journal of the IGPL*.

[Hansen,2002] H. H. Hansen. The straw thing of fallacy theory: the standard definition of "fallacy". *Argumentation*, 16(2):133–155, 2002.

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

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

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

[Kirsh and Maglio, 1994] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18:513–549, 1994.

[Magnani, 2001a] L. Magnani. *Philosophy and Geometry. Theoretical and Historical Issues*. Kluwer Academic Publisher, Dordrecht, 2001.

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

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

[Magnani, 2005] L. Magnani. Creativity and the disembodiment of mind. In P. Gerv´as, A. Pease, and T. Veale, editors, *Proceedings of CC05, Computational Creativity Workshop, IJCAIO2005, Edinburgh*, pages 60–67, 2005.

[Magnani, 2006a] L. Magnani, editor. *Abduction, Practical Reasoning, and Creative Inferences in Science*, 2006. Special Issue of the *Logic Journal of the IGPL*.

[Magnani, 2006b] L. Magnani. *Knowledge as a Duty. Distributed Morality in a Technological World*. Cambridge University Press, Cambridge, 2006.

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

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

[Peirce, 1931–1958] C. S. Peirce. *Collected Papers*. Harvard University Press, Cambridge, MA, 1931–1958. 1–6, ed. by C. Hartshorne and P. Weiss, 7–8, ed. by A. W. Burks, 1931-1958.

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

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

[Thagard, 2000] P. Thagard. *Coherence in Thought and in Action*. The MIT Press, Cambridge, MA, 2000.

[Woods, 2004] J. Woods. *The Death of Argument*. Kluwer Academic Publishers, Dordrecht, 2004.

[Woods, 2005] J. Woods. Epistemic bubbles. In S. Artemov, H. Barringer, A. Garcez, L. Lamb, and J. Woods, editors, *We Will Show Them: Essays in Honor of Dov Gabbay*, volume II, pages 731–774. College Publications, London, 2005.

Abductive Issues in Peirce's Proof of Pragmaticism

Ahti-Veikko Pietarinen
Department of Philosophy
University of Helsinki
ahti-veikko.pietarinen@helsinki.fi

1. Introduction

Abduction and Pragmaticism

Our journey begins with Charles S. Peirce's later writings on his philosophy of pragmaticism. He set out developing his mature views around the mid-1890s and named his position in 1903. My purpose is not to delve into all of the intricate details of the development of that thought over the last years of his life, but to argue for three interrelated theses: (i) that Peirce believed his pragmaticism to incorporate the notion of abduction, (ii) that pragmaticism is a consequence of several fundamental assumptions that he believed can be shown to build a logically sound argument for its correctness, and (iii) that how abduction features in pragmaticism is that it contributes to the argument in terms of two fundamental steps towards its conclusion.

Accordingly, I do not intend to discuss the broad questions of what the nature of abduction is or should be, or how it is applied across the broad scenery of scientific inquiry. Nonetheless, the argument that I am essaying here does, I believe, constitute some answers towards a clarification of the nature of abduction (or the nature of abductive inference, if it is inference), but I believe it to be prudent to first clarify the role abduction plays in the larger framework of Peirce's pragmaticism. It also appears that this larger framework has been sadly neglected in contemporary discussions concerning Peirce's notion of abduction.

The key passage that I am alluding to is taken from Peirce's *Harvard Lectures* delivered off-campus in 1903. There he states his well-known belief that pragmaticism and abduction are two sides of the same philosophical thought. In particular, he believes that abduction is subsumed by pragmaticism in the sense that the logic of abduction does not answer any questions which the Maxim of Pragmaticism, upon which pragmaticism relies, is not geared to answer:[1]

> If you carefully consider the question of pragmatism you will see that it is nothing else than the question of the logic of abduction. … The maxim of pragmatism, if true, fully *covers* the entire logic of abduction. (CP 5.196, 1903, *Harvard Lectures VII*; PPM: 249)

The key term in this passage is "covers". Peirce appears to think that as soon as the Maxim of Pragmaticism is approved as being correct, we will *ipso facto* have an approval of the logic of abduction.

Even more stoutly, he states in the same lecture that the Maxim of Pragmaticism *is*, in fact, the logic of abduction:

> The maxim of Pragmatism, if it is sound, or whatever ought to replace it, if it is not sound, is nothing else than the logic of abduction. (PPM: 282-283, 1903, *Harvard Lectures VII*)

Thus we had better look more closely at what the maxim is, how its formulations developed over time and what its ultimate form was upon which Peirce thought to have finally obtained a logically sound proof for its correctness.

The intense relationship between pragmaticism and abduction needs closer scrutiny in Peirce scholarship than what has been provided thus far. Neglecting this issue will prevent us from gaining true insights into the nature of the logic of abduction.

1 I am using the term "Maxim of Pragmaticism" to refer to the formulations of the maxim that Peirce often termed the "Maxim of Pragmatism" or the "Pragmatic Maxim". All references pertain to Peirce's later philosophy, which he termed pragmaticism but did not always use that ugly term consistently, and not to his earlier pragmatism which he began developing in the late 1860s. Apel (1995) and Pietarinen & Snellman (2006) discuss the fundamental differences between the two philosophies.

The Maxim of Pragmaticism

What is the conception of the Maxim of Pragmaticism that Peirce gravitated to in his later philosophy? Several formulations with some negligible differences are found in his vast corpus. What is common to all is the purpose that the maxim is set out to accomplish. It is the same with his earlier pragmatism, to achieve the "third grade of clearness" (CP 5.402, 1878, *How to Make Our Ideas Clear*) for the apprehension of intellectual concepts. By intellectual concepts, he means those that can contribute to the validity of arguments and are such that bear a truthful relationship with their objects.

One concise formulation of the maxim is the following: "If one can define accurately all the conceivable experimental phenomena which the affirmation or denial of a concept could imply, one will have a complete definition of the concept, and there is absolutely nothing more in it" (CP 5.412, 1905, *What Pragmatism Is*). The conceivable phenomena of the effects of an object of conception are its "would bes" or "may bes". The meaning is the relationship or the form that obtains between the sign and all of its possible or conceivable implications. These relationships or forms are real generals external to linguistic or mental realms of human beings.

Another formulation originating around the same time states that "the entire intellectual purport of any symbol consists in the total of all general modes of rational conduct which, conditionally upon all the possible different circumstances and desires, would ensue upon the acceptance of the symbol" (CP 5.439, 1905, *Issues of Pragmaticism*). Although Peirce here uses symbols as the subject matter to which the maxim can be applied, some alternative formulations are quite clear on the point that they are not meant to rule out icons and indices as intellectual signs that fall within the domain of pragmatic meaning.

At any event, the method of pragmaticism is a method of analysis of pragmatic meaning. It is intended to provide us with the "third grade of clearness" of intellectual concepts. This is facilitated through an investigation of what the relationships are that signs have to their various interpretants.

Pragmaticism is not a system or doctrine of philosophy but a method of ascertaining how our thinking and living contributes to meaning. According to Peirce's thought, expressed very early on in his career, we should not "pretend to doubt in philosophy what we do not doubt in our hearts" (W 2:212, 1868, *Some Consequences of Four Incapacities Claimed for Man*). This means that any proof of the correctness of the maxim has to be prepared to account for how our common human activities of getting across the meaning of intellectual signs are linked with the conception of what it means that such signs can be true as well as that they can be false across all contexts in which they may be encountered and contemplated.

The purpose of the pragmatistic method is thus to reveal what the logical structure of meaning is. Peirce took semeiotic, or the theory of signs, to be formal semiotics, or more accurately speaking, that study of what we nowadays call logic or formal semantics. What the proof aims at is an account of meaning that shows the form or the relational structure produced by general, interactive behaviour and action across all conceivable contexts. The meaning of concepts is to reveal practical outcomes by exploring and experimenting upon conditional hypotheses that assert the possibility of those contexts.

In the context of scientific inquiry and discovery, before applying the scientific method to any particular hypothesis to test its truth by induction, pragmaticism asserts that we must already have conceived of the meaningfulness of the propositions that assert that claim. The way to assess meaningfulness is to observe and contemplate the conceivable outcomes of the applications of the concepts in question. Therefore, applying the Maxim is the initial step to be performed when applying the scientific method to any phenomenon.

2. The Proof of Pragmaticism

History

What is particularly interesting is that Peirce claimed to have "an ample supply of seductive persuasions to pragmatism" and "two or three scien-

tific proofs of its truth" (CP 5.468, 1907, *A Survey of Pragmaticism*, cf. Fisch 1986, Robin 1997). Moreover, he thought such a proof inexorably hard to produce:

> If I may trust my most cautious logic, such a proof I have worked out and perfected. "What is it, then? Produce it." Unfortunately, like many another intricate proof[s], it only becomes evident upon close, severe, and long study. (MS 322, February/March 1907, *Pragmatism*)

A couple of years after Peirce had outlined his project of pragmaticism and argued for its similarity to the logic of abduction, he came to believe that he had solved the problem of its proof. This was achieved by replacing, as he states, "by a scientific and logical proof the merely rhetorical defence I made of the principle in my two original articles in the Popular Science Monthly of Nov. 1877 and Jan. 1878, such being the only sort of argument which would be at all admissible in a popular journal" (MS 296, March 1908, *The First Part of an Apology for Pragmaticism*, date according to Max Fisch's notes). A few months earlier, Peirce had remarked to his friend Francis C. Russell that "the article to be written is the most difficult to present of any that I have ever written. It is the first part,--the lemma,--to my proof of the truth of pragmaticism" (MS L 387: 3, 7 December 1907).

What is the proof? Can it be explained or, better, reconstructed? Where does the notion of abduction surface in his proof? Before answering these, we must dwell upon one more key concept of Peirce's pragmaticism.

Habits

According to Peirce, a habit is a "generalizing tendency" (CP 6.204, 1898, *The Logic of Continuity*). It is not simply a conscious operation in the human mind. Habits are real generals found in nature rather than in the human psyche:

> Let us use the word 'habit' not in its narrower, and more proper sense, in which it is opposed to a natural disposition (for the term *acquired habit* will perfectly express that narrower sense), but in its wider … sense, in which it denotes such a specialization, original or acquired, of the nature of a man, or an animal, or a vine, or a crystallizable chemical substance, or anything else, that

> he or it will behave, or always tend to behave, in a way describable in general terms upon every occasion. … that may present itself of a generally describable character. (CP 5.538, c.1902, *Belief and Judgment*)

"From the human mind", Peirce states, "we only collect external information about habit. Our knowledge of its inner nature must come to us from logic" (NEM 4:142). The logic that provides us insights into the inner nature of habits is the same logic that guides the actions that we need to take whenever new situations arise that need skill and planning for optimal decision making.

Moreover, habits do not reduce to singular actions or sequences of action, but pertain to the same category as general recommendations or plans that articulate general directions of action that goal-directed tasks of acting for a purpose have.

> It would be necessary, in order to define a man's habit, to describe how it would lead him to behave and upon what sort of occasion---albeit this statement would by no means imply that the habit *consists* in that action. (CP 2.664, 1878, *The Doctrine of Chances*, corrections in 1910)

The logic that provides us with the knowledge of the inner nature of habits is the logic of abduction. Essential to it is the experimentation that we perform in the mind with signs that are liable to provide the new information and new insight assisting us in formulating the general recommendations and plans essential in decision-making tasks. Such signs are typically diagrammatic, and the most common diagrammatic signs are visual. Visual diagrams are signs that are cognitively the most economic and the most fertile for application in thought experiments. According to Peirce, every experiment is an operation of thought:

> The [logical] activity takes the form of experimentation in the inner world; and the conclusion … is that under given conditions, the interpreter will have formed the habit of acting in a given way whenever he may desire a given kind of result. (CP 5.491, 1907, *A Survey of Pragmaticism*)

This paragraph comes close to the conclusion Peirce sets out to demonstrate by his "seductive persuasions" for the truth of the Maxim of

Pragmaticism. It is by abductive reasoning that a conclusion of acting in a given way for a given kind of result is arrived at. Such a conclusion is the habit that agents appeal to whenever acting for a purpose. Similar acts of discovery take place in scientific reasoning processes, where the outcome will be the formation of new scientific hypotheses.

Habits and Pragmaticism

Habits are real generals out there in the wild. They are relational mappings from all possible and conceivable situations to actions. These relations, which by no means are invariably salient to human beings, need to be searched for, discovered, pointed out, hit upon and mediated by something. The role of these tasks is fulfilled by signs.

Peirce speaks of a habit of making good guesses, or conjectures, which thus make them not just blind guesses or some inexplicable powers for hypothesis generation. Thus the intellectual guesswork that we do in science is not blind because we have epistemic access to those habits. We also have a mechanism for changing or improving on them at hand whenever such occasions arise:

> The essential function of a sign is to render inefficient relations efficient, … not to set them into action, but to establish a habit or *general rule* whereby they will act on occasion. ... [A] sign is something by knowing which we know something more. (CP 8.332, 1904, *Letter to Lady Welby*)

Signs themselves do not map situations into actions, but they do provide vital guidelines to equip us with a diversity of methods of searching for information which facilitates those processes that give us objects of signs in all situations consistent with what evidence we have.

As soon as the objects of signs are found, the interpreter can proceed further in the interpretation of signs. In particular, there are many interpretations that are logical and rely on self-controlled processes. Pragmaticism concerns the establishment of habits as such logical interpretants that emerge in self-controlled action:

> Under given conditions, the interpreter will have formed the habit of acting in a given way whenever he may desire a given kind of result. The real and living logical conclusion *is* that habit; the verbal formulation merely expresses it. … But *action* cannot be a logical interpretant, because it *lacks generality*. … Consequently, the most perfect account of a concept that words can convey will consist in a *description of the habit* which that concept is calculated to produce. But how otherwise can a habit be described than by a *description of the kind of action to which it gives rise*, with the specification of the conditions and of the motive? (CP 5.491, 1907, *Pragmaticism*; MS 318; EP 2:418)

Again, we have come close to the conclusion that expresses the maxim. The relationship between habits and the maxim is that habits, which the agents have acquired from experience and action, now become the beliefs of agents, and refer to those circumstances, actual and possible alike, upon which agents are prepared to act on any occasion. Just as possible scenarios or figments of worlds represent states of affairs for which we must be prepared, so do habits, which become grown and unravelled as we explore the contingencies, the hypothetical states of affairs that constitute the practical bearings or conceivable effects of the concepts we strive to understand and evaluate.

From Habits to Strategies

At this point we arrive at the junction at which Peirce's pragmaticism merges with much more recent decision-theoretic and game-theoretic ways of thinking and acting, in that habits are seen to be closely related to strategic behaviour and action (Pietarinen 2006). Note the force of the following fragment snipped from the previous passage:

> Action cannot be a logical interpretant, because it lacks generality … But how otherwise can a habit be described than by a description of the kind of action to which it gives rise, with the specification of the conditions and of the motive? (CP 5.491, 1907, *Pragmatism*; MS 318; EP 2)

Likewise in game theory, strategies are identified through the individual actions they give rise to. In game theory, strategic rules are general and can only be described or explicated with reference to particular actions, but can-

not be reduced to them. Only terminal points are assigned with payoffs, in terms of which equilibria are then defined. No individual set of actions is in itself representative of what the future courses of actions in any conceivable circumstance ought to be. Only a completely general strategy, a function from situations to actions, can tell that.

What is more, strategies that are to be of any real value need to be defined not only in view of probable situations, but also in view of those that lie on off-equilibrium paths:

> Now, the identity of a habit depends on how it might lead us to act, not merely under such circumstances as are likely to arise, but under such as might possibly occur, no matter how improbable they may be (No matter if contrary to all previous experience [Marginal note, 1893]). (CP 5.400)

Peirce's notion of a habit as a general plan or recommendation of action thus comes very close to what nowadays is meant by a strategy in game theory.[2]

The Proof of Pragmaticism

Peirce's Game-Theoretic Semantics

Peirce's pragmaticism can be regarded at heart as a theory for which his system of signs does not yield deep enough explanatory insight. Rather, it is a theory about formal semeiotic, or a theory about logical semantics, and is to be studied as such. What he has to say about different kinds of signs may be attractive in its own right, but the various classificatory schemes that he manufactured in his spare time are not likely to improve our understanding of his logical philosophy and the theory of meaning, at least not without relating his remarks to developments in various areas of contemporary logic (Hilpinen 2004, Houser et al. 1997, Pietarinen 2006). For Peirce, logic *is* formal semeiotic. It concerns the logical analysis of intellectual signs.

Indeed, Hilpinen (1982) showed that in his theory of the meaning of intellectual signs, Peirce anticipated Game-Theoretic Semantics (GTS, see also

2 Hintikka (1999, p. 100) notes that, "I suspect, in other words, that inside each Peircean habit there lurks (at least in the area of epistemology) a strategic rule trying to get out". See also Kilpinen (2000) and Pietarinen (2006).

Hintikka 1999, Pietarinen 2006). That suggestion can now be carried even further in the light of evidence from unpublished manuscripts. In some of them, those anticipations are strikingly palpable. Consider, for instance, the following passage from a 1905 manuscript *The Basis of Pragmaticism*:

> A critical analysis of the nature of a sign would show that the action requires a source of concepts to be conveyed, and therefore *in some sense a mind* from which the concepts, propositions, and arguments are conveyed to the mind of the interpreter; and *the two minds* must be *capable of coming to an understanding and of observing it* when it is reached. This supposes *a power of deliberate self-controlled thinking*. Now nothing can be controlled that cannot be observed while it is in action. It is therefore requisite that both minds but especially the [Utterer-mind] should have *a power of self-observation*. Moreover, control supposes a capacity in that which is to be controlled of acting in accordance with *definite tendencies of a tolerably stable nature*, which implies a reality in this governing principle. But these habits, so to call them, must be capable of being modified according to some ideal in the mind of the controlling agent; and this controlling agent is to be the very same as the agent controlled; the control extending even to the modes of control themselves, since we suppose that the interpreter[-mind] under the guidance of the [Utterer-mind] discusses the rationale of logic itself. (MS 280: 30-32, 1905, *The Basis of Pragmaticism*, added emphasis)

The relationship between Peirce's theory of pragmaticism and GTS is as follows. What "the two minds" in this quotation refer to are the Utterer and the Interpreter, or the Verifier and the Falsifier, who in GTS undertake to show the material truth or the falsity of the given assertion. What "capable of coming to an understanding and of observing it" means is that the payoffs are determined at the terminal histories of any play of the game. The "power of deliberate self-controlled thinking" and the "power of self-observation" can be correlated with strategic thinking and planning. Finally, Peirce's notion of self-control together with the "definite tendencies of a tolerably stable nature" can be correlated with the existence of certain winning strategies, or habits of acting for a purpose, or any notion of equilibrium as it is used in game theory associated with acting for a purpose.

The Proof: Background

What, then, is Peirce's much advertised proof of pragmaticism? Can it be reconstructed given the conceptual machinery that we possess? It turns out that Peirce's proof can be related to the verificationist game-theoretic conception of meaning (Pietarinen & Snellman 2006).

At the bottom lies the Hintikka-type outlook of logic as formal semeiotic and a model-theoretic and calculistic analysis, which consists of three tenets: (i) Applications of game theory provide a way to understand the notion of truth. (ii) Language is a re-interpretable system. (iii) Semantics is effable, in other words the relationship between language and the world can be articulated and theorised about by the use and application of language.

The analysis is model theoretic, since the purpose of Peirce's version of GTS was to link assertions with the world to its universe of discourse and to provide conditions for their truth. It is calculistic, because new systems, methods and logics are created as the need arises and in the course of fulfilling the purpose of the inquiry.

The Proof

What Peirce wished to demonstrate in 1907 is his "Kernel of Pragmaticism":[3]

> The *total* meaning of the predication of an intellectual concept consists in affirming that, under all conceivable circumstances of a given kind, the subject of the predication would (or would not) behave in a certain way,---that is, that it either would, or would not, be true that under given experiential circumstances … certain facts would exist. (EP 2:402, 1907, *Pragmatism*)

We proceed with the reconstruction of the proof of this "kernel" as follows:

1. Every thought, intellectual concept and generality (**tig**) is a sign.

a. If thinking is dialogue between the utterer and the interpreter, every **tig** is a sign.

b. Thinking is dialogue between the utterer and the interpreter.

3 See the textual evidence of MS 318 reprinted, in part, in EP 2.

2. Something is a **tig** sign, iff it is possible for it to have the essential ingredients (*quaesitum*) of the utterer and the interpreter.

3. A **tig** sign can have the essential ingredients of the utterer and the interpreter (*quaesitum*), iff it can have an object (*requaesitum*).

4. A **tig** sign has an object, iff the individuals that constitute the extension of the **tig** sign can be picked in a semantic game (whose strategies are winning or losing on the grounds that take epistemic factors into account).

5. The individuals constituting the extension of the **tig** sign can be picked in a semantic game only if the sign has an interpretant.

6. → Hence, by definition something is a **tig** sign, iff it mediates between the object and the interpretant and it has an interpretant determined by the object and the object determines it according to the interpretant.

*7. All **tig** signs give their objects as answers to questions that can be answered by them.*

8. We can acquire conclusive answers (i.e., an object of a **tig** *sign) to our questions only by making experiments on various ways of finding solutions in our thoughts.*

9. All **tig** signs give their objects through habits that we have received through experiments made in our thoughts.

10. → Final logical interpretants of **tig** signs are habits that we have received through experiments made in our thoughts.

Such is the argument for the truth of the pragmaticism's kernel. Its paraphrase is that the whole meaning of a **tig** sign is that "certain kinds of events would happen, once in so often, in the course of experience, under certain kinds of existential circumstances" (EP 2:402, 1907, *Pragmatism*). In other words, the final logical interpretants of **tig** signs, their meanings, are habits of acting in a certain way in certain circumstances, which is to say are mappings from situations to actions.

Peirce's interpretation of assertions and logical constants is virtually that of GTS. If the interpreter believes two assertions, he is entitled to choose the weaker, and in interpreting quantifiers, he may choose an individual to be the interpretation of a universal assertion, and the utterer may choose an individual to be the interpretation of an existential assertion.

Peirce's views on beliefs, thinking and truth are considerably clarified against the backdrop of GTS. To have a belief is to possess a habit of acting in a certain way in different circumstances, not necessarily in the actual circumstances but in those which refer to the future, while logical constants contribute to that habit by making the selections of assertions and individuals possible.[4] This is "the specification of the conditions", or the description of the actions, which in turn is what habits are geared to describe (CP 5.491; MS 318; EP 2:418). The specification of the "motive" (CP 5.491; MS 318; EP 2:418) is in the activities of the verification and falsification of assertions. What belief means is that what would happen if we were to act according to habit, not in any particular situation, but in all possible situations.

3. Conclusions

Abduction in the Proof of Pragmaticism

By 1907, Peirce had relocated the place of the logic of abduction under his pragmaticism, rather than the two being coextensive in the sense of the one being "nothing else" (CP 5.196; PPM: 249) than the other. Our attention turns to two clauses in the proof, clauses 7 and 8. Together with the definition of the sign in clause 6 (which we might term the 'lemma' of the proof) and with the clause 9 they yield the desired conclusion.

Clauses 7 ("All **tig** signs give their objects as answers to questions that can be answered by them") and 8 ("We can acquire conclusive answers [an object of the **tig** sign] to our questions only by making experiments on various ways of finding solutions in our thoughts") indeed have a close connection with abduction. Abductions, of abductive 'inferences', are answers to the questions that the inquirer puts to the (definite, reliable) sources of information. This causes this form of reasoning to be entirely different from both deduction and induction.

> By its very definition abduction leads to a hypothesis which is entirely foreign to the data. To assert the truth of its conclusion ever so dubiously

4 "The species of future tense of the logical interpretant is that of the conditional mood, the '*would be*.' … The logical interpretant should in all cases be a conditional future" (EP 2:410).

> would be too much. There is no warrant for doing more than putting it as an interrogation. To do that would seem to be innocent; yet if the interrogation means anything, it means that the hypothesis is to be tested. Now testing by experiment is a very expensive business, involving great outlay of money, time, and energy; so that comparatively few hypotheses can be tested. Thus, even the admission of an abductive conclusion to the rank of an active interrogation is a concession not to be too lightly accorded. (MS 692, 1901, *The Proper Treatment of Hypotheses: a Preliminary Chapter, toward an Examination of Hume's Argument against Miracles, in its Logic and in its History*; HP 2:898-899)

How answers to questions are obtained is by having the general forms of habits at our disposal and by being able to make the needed thought experiments. Recalling that thought experiments are for Peirce a species of scientific experiments, this is really what clause 9 in the proof asserts. And it constitutes the final step towards the conclusion.

Abduction and Interrogation

In Peirce's proof of pragmaticism, information-seeking activities make answers to questions possible. In turn, the information received brings about those activities of seeking and finding that contribute to the meaning of signs. The purpose is to produce objects of signs in all situations consistent with the given evidence.

The strategies of seeking and finding are the final logical interpretants, in other words habits. Every experiment is an operation of thought, and so experimenting upon such strategies is to receive information from the source.

Justification of the conclusion produced by abduction is only received through certain methods of interrogation and constructions of question-answer sequences by which we find the objects of signs. This is what Hintikka (1999) has in so many words emphasised to be the characteristic feature of Peircean abduction and which is the kind of reasoning whose "legitimacy depends upon altogether different principles from those of" deduction and induction (CP 6.525, 1901, *Hume on Miracles*).

Interrogation may also be part of the process of arriving at some explanatory hypothesis. Therefore, it is in operation both in the scientific pro-

cesses of discovery and in the scientific processes of justification. The latter involve those of hypothesis selection. If testing is required, the process leaps into the realm of inductive reasoning and testing by induction.

Pragmaticism and Dialogues

The purpose of these interrogatory and interactive actions is to hoard information from a source that authorises suitable objects to be picked to satisfy intellectual signs. The meaning of a sign is in the habits that shape such structures and possibly change them, in which case the meanings will change. We may spell out the logical structure of pragmatic meaning in terms of the relational structure of dialogical interaction that the application of habits gives rise to.

In other words, dialogues between the utterer and the interpreter (or the agent and the environment) produce a geometry which is the meaning of the assertion. This relational structure depicts all conceivable effects (payoff distributions at terminal histories) that an intellectual concept may possibly have. The existence of habits (winning strategies or some stable strategy profiles) agrees with the truth of assertions. Recall here Peirce's conclusion concerning his account of pragmaticism:

> The most perfect account of a concept that words can convey will consist in a description of that habit which [the concept] is calculated to produce. (MS 318, 1907, *Pragmatism*)

Outside of philosophy, similar concepts of meaning have become influential in, for instance, computational and logical theories of interaction and communication.

Acknowledgements

This work has been supported by the three-year Excellence in Research Grant by the University of Helsinki (*Peirce's Pragmatistic Philosophy and Its Applications*, Principal Investigator Ahti-Veikko Pietarinen). My warmest thanks go to the organisers and the participants of the 2006 *Lisbon International Meeting on Abduction and the Process of Scientific Discovery*.

References

Apel, Karl-Otto 1995. *Charles S. Peirce: From Pragmatism to Pragmaticism*, New York: Humanity Books.

Fisch, Max 1986. The 'proof' of pragmatism, in Max Fisch, *Peirce, Semeiotic, and Pragmatism: Essays by Max H. Fisch*. Bloomington: Indiana University Press, 362-375.

Hilpinen, Risto 1982. On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics, *The Monist* 65, 182-188.

Hilpinen, Risto 2004. Peirce's logic, in Dov M. Gabbay and John Woods (eds), *Handbook of the History and Philosophy of Logic* 3, Amsterdam: Elsevier.

Hintikka, Jaakko 1999. What is abduction? The fundamental problem of contemporary epistemology, in Jaakko Hintikka, *Inquiry as Inquiry*, Dordrecht: Kluwer, 91-113.

Houser, Nathan, Roberts, Don D. and Van Evra, James (eds), 1997. *Studies in the Logic of Charles Sanders Peirce*, Bloomington: Indiana University Press.

Kilpinen, Erkki 2000. *The Enormous Fly-wheel of Society: Pragmatism's Habitual Conception of Action and Social Theory*, doctoral dissertation, University of Helsinki.

Peirce, Charles S. 1931-58. (CP) *Collected Papers of Charles Sanders Peirce*, 8 vols., ed. by Charles Hartshorne, Paul Weiss, and Arthur W. Burks. Cambridge, Mass.: Harvard University Press.

Peirce, Charles S. 1967. (MS) Manuscripts in the Houghton Library of Harvard University, as identified by Richard Robin, *Annotated Catalogue of the Papers of Charles S. Peirce* (Amherst: University of Massachusetts Press, 1967), and in The Peirce Papers: A supplementary catalogue, *Transactions of the C. S. Peirce Society* 7 (1971), 37-57.

Peirce, Charles S. 1976. (NEM) *The New Elements of Mathematics*, 4 vols., ed. by Carolyn Eisele. Berlin: Mouton; Atlantic Highlands, NJ: Humanities Press.

Peirce, Charles S. 1980-. (W) *Writings of Charles S. Peirce: A Chronological Edition*, Peirce Edition Project, Bloomington: Indiana University Press.

Peirce, Charles S. 1985. (HP) *Historical Perspectives on Peirce's Logic of Science. A History of Science*, 2 vols, Ed. by Carolyn Eisele. Berlin: Mouton Publishers.

Peirce, Charles S. 1997. (PPM) *Pragmatism as a Principle and Method of Right Thinking. The 1903 Harvard Lectures on Pragmatism by Charles S. Peirce*, ed. by Patricia Ann Turrisi. Albany: State University of New York Press.

Peirce, Charles S. 1998. (EP) *The Essential Peirce* 2, Peirce Edition Project, Bloomington: Indiana University Press.

Pietarinen, Ahti-Veikko 2006. *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication* (Synthese Library 329), Dordrecht: Springer.

Pietarinen, Ahti-Veikko and Lauri Snellman 2006. On Peirce's late proof of pragmaticism, in Tuomo Aho and Ahti-Veikko Pietarinen (eds), *Truth and Games*, Helsinki: Acta Philosophica Fennica, 275-288.

Robin, Richard S. 1997. Classical pragmatism and pragmatism's proof, in Jacqueline Brunning and Paul Forster (eds), *The Rule of Reason: The Philosophy of Charles Sanders Peirce*. Toronto: University of Toronto Press, 139-152.

Abduction's dark ways

W.F.G. Haselager
Nijmegen Institute for Cognition and Information
Radboud University
w.haselager@nici.ru.nl

"*If we want to say that logic deals with thinking, we had better say that logic teaches us how thinking should proceed and not how it does proceed. (...) We know very well that productive thinking is bound to follow its own dark ways, and that efficiency cannot be secured by prescriptions controlling the paths from the known to the unknown. It is rather the results of thinking, not the thinking processes themselves, that are controlled by logic.*" Reichenbach (1947, p.1).

"*Abduction really is a terrible problem for cognitive science, one that is unlikely to be solved by any kind of theory we have heard of so far.*" Fodor (2000, p. 41).

Introduction

Traditionally, abduction is characterized as a two-step process: the generation of a set of explanatory hypotheses, followed by the selection of the 'right' (most probable, reasonable, justified, etc.) one. In cognitive science, the inability to understand how plausible explanatory hypotheses get generated has hampered progress in the modeling of common sense reasoning and behavior. Human beings have an amazing ability to jump from a few observations to an understanding of what is going on in their environment and act on the basis of that. Upon hearing a strange sound in the kitchen, one may think that one forgot to turn off the coffee machine and then do so after all. Humans, and, of course, animals in their own natural environment too, 'get it' a lot if not most of the time, and quite fast too. A guiding assumption of this paper is that 'getting it' is akin to the generation of an explanatory hypothesis that explains the events witnessed, thus providing a

frame for relevant action. However, it has proven to be very difficult indeed within cognitive science to model the ability to generate plausible hypotheses about what is going on, such that appropriate behavior can be displayed. So far, the traditional approach based on purely abstract internal knowledge representation has failed to make substantial progress in modeling abduction under realistic circumstances (Pylyshyn, 1987; Haselager, 1997; Fodor, 2000). Thus, abduction, as a main mode of productive thinking, not only seems to be 'dark' from a logical point of view, as Reichenbach indicates, but also turns out to be, in Fodor's words, a 'terrible problem' from the perspective of psychology and cognitive science. Regarding the latter issue, I propose that a first step towards elucidation may lie in considering cognition and inference from a more encompassing perspective than the traditional view allows. This proposal is inspired by two observations: First, Peirce (CP 5.181) has suggested that perception and inference should be seen as lying on a continuum rather than as constituting a dichotomy. Secondly, several experiments in developmental psychology seem to demonstrate a strong perceptual influence on so-called 'higher' cognitive processes, such as abductions regarding causality and animacy. These experiments are in line with an increasing attention within cognitive science for the embodied embedded nature of cognitive processes. According to this more recent view, a lot of our thought is taking place not separated from, but intermingled with the way we perceive and do things. Like other cognitive processes, abduction may not be separable from the perceptual-motor interaction between the system and its world. Adopting this more encompassing perspective on abductive inference may even help to elucidate several aspects of scientific practice and methodology as well.

Abduction and cognitive science

The problem of abduction has played a central, if not always fully recognized, role in cognitive science, specifically in the attempt to understand common sense. Traditionally, a distinction is made between cognitive modules and a central system (Fodor, 1983). Modules are processing systems that are domain-specific (e.g. vision), operate very fast and in a mandatory

way, and are informationally encapsulated, in the sense that they use only limited information for their processing. In contrast, the central system underlies our 'higher' cognitive processes. It is responsible for creating and updating a world model and keeping it consistent. It does this by efficiently using relevant parts of our common sense knowledge in order to 'make sense' of the world. As Fodor (1983, p.114) put it, it answers the following question: "What, in general, is the optimal adjustment of my beliefs to my experiences?"

The problem in computationally modeling the answer to this question is basically twofold. First of all, the optimal adjustment of beliefs to new experiences depends in large part on what one already knows. But human beings know *a lot*. For instance, try to describe what one has to know in order to understand 'having a beer together'. Not only does this require lots of biological and physical knowledge about glasses or bottles, liquids, gravity, drinking, thirst, effects of alcohol, etc., but also about social factors such as friendship or politeness, appropriate places and circumstances for drinking, etc. Attempts within Artificial Intelligence to specify everything we know (Guha & Lenat 1990, 1994; Singh & Minsky, 2002; Singh, et al. 2004) have not met with clear success, despite considerable efforts. A second and related problem concerns the *efficient use* of common sense knowledge. Even if one could represent all common sense knowledge in sufficient detail, how would a computational system find the *relevant* parts of the immense database within a very short amount of time? This problem is aggravated by the lack of understanding of how 'relevance' is determined. It seems clear that context is of paramount importance, but how is not well understood.

Thagard's (1988) computational model PI (for 'Processes of Induction') provides an early and influential example of the way symbolic AI has tried to deal with abductive processes. PI uses frames, well-structured representational formats with default values for typical features, to represent concepts, locating them within conceptual hierarchies. The basic mechanism for modeling a search for relevant information consists of the directed spreading of

activation through the frames, their incorporated rules and the structure of the conceptual hierarchy:

> "The system is computationally efficient because only plausibly relevant rules - those attached to active concepts - are likely to be considered on a processing cycle." Holland, et al. (1986, p.132; see also Thagard, 1988, p.20-22, 54).

Models such as PI are fairly typical of AI approaches to abduction. They focus on 'pure thought': abstract (propositional) internal information processing, without any direct connection to perception or motor activities. These models however, have run into the problem of scaling up to a realistic level of complexity. Thagard (1988, p.29) mentions about 60 rules and concepts as an upper limit for PI. This is not just a specific feature of traditional symbolic models. The same problem besets connectionist models. Thagard later developed a connectionist implementation (called ECHO: 'Explanatory Coherence by Harmany Optimization') of the conceptual frameworks relevant to various scientific discoveries, and says about his connectionist model in relation to Lavoisier's oxygen theory:

> "A full computational analysis of Lavoisier's mental system would specify all its conceptual and propositional components and clarify their interrelationships, but a cognitive system rich enough to do this and incorporate ECHO has yet to be built." Thagard (1992, p.84).

It is important to realize that this problem has not just to do with lack of time or computational resources. The problem is more principled in that there is no clear understanding yet of what knowledge needs to be represented in such a way that the processes operating upon it can quickly find the relevant parts of it. These problems show themselves vividly, as soon as a, in fact rather low, threshold of scale (i.e. the number and structure of knowledge items involved) is surpassed for classical symbol-rule and connectionist models alike (see Haselager & van Rappard, 1998 for more details on connectionism). For this reason, Putnam (1992, p.8-14), for one, has argued that the whole AI-enterprise is 'much ado about not very much', while Fodor (1983, p.107) has spoken of 'Fodor's first law of the non-existence of

cognitive science'. According to both philosophers, certain key ideas in logic and the philosophy of science concerning hypothesis generation and confirmation have to be developed before cognitive science and AI can model central processes successfully. I propose to follow a different track however, by steering away from abstract purely internal information processing and instead investigating the influence of sensorimotor processes on abduction.

Abduction shading into perception

As is well known, Peirce claimed that abduction and perception should not be seen as a dichotomy:

> "*Abductive inference shades into perceptual judgment without any sharp line of demarcation between them*; or, in other words, our first premisses, the perceptual judgments, are to be regarded as an extreme case of abductive inferences, from which they differ in being absolutely beyond criticism. The abductive suggestion comes to us *like a flash*. It is an act of insight, although of extremely fallible insight. It is true that the different elements of the hypothesis were in our minds before; but it is the idea of putting together what we had never before dreamed of putting together which flashes the new suggestion before our contemplation. On its side, the perceptive judgment is the result of a process, although of a process not sufficiently conscious to be controlled, or, to state it more truly, *not controllable and therefore not fully conscious*." Peirce (CP 5.181; my emphasis).

The continuity between perception and abduction, as envisioned by Peirce, stands in sharp contrast with the view of abduction as a separate, abstract, process performed by a central cognitive system. In his famous work on discovery, Hanson (1958, p. 86) explicated this by taking the phrase 'seeing that' quite literally:

> "Peirce regards an abductive inference (..) and a perceptual judgment as being opposite sides of the same epistemological coin. *Seeing that* is relevant here. (...) The dawning of an aspect and the dawning of an explanation both suggest what to look for next. In both, the elements of inquiry coagulate into an intelligible pattern. The affinities between seeing the hidden man in a cluster of dots and seeing the Martial ellipse in a cluster of data are profound."

In the 90's an approach to cognition and behavior developed that emphasized the importance of bodily interaction with the environment for a proper analysis of cognition and intelligent behavior, which I will refer to here as Embodied Embedded Cognition (EEC).[1] According to this view it makes no sense to posit the existence of a central system that can be studied on its own. First of all, what traditionally is called 'central' or 'higher' cognition is thoroughly and recurrently connected with so-called 'peripheral' processes of perception and action. Secondly, these sensorimotor processes cannot be completely understood without a proper understanding of the intrinsic dynamics (Kelso, 1995) of the physical system itself, i.e. of the morphology of the body, and of the way this facilitates or hampers specific coordination tendencies. For instance, the spring-like quality of the legs of human beings provides specific constraints (and opportunities) for controlling leg movements. Understanding the task of the motor cortex requires a consideration of these constraints. Thirdly, the environment of the organism is considered an integral part of the organism's cognition. Organisms engage in a continuing series of interaction cycles with the environment, in which behavior evokes a behavioral response that will help to shape the organisms' behavior at the next instance. Studying a central system on its own may be as misleading as trying to study a leg without realizing that it is accompanied by one or three other legs, supporting the rest of the body and allowing its movement in a balanced way on a variety of surfaces etc. The resulting analysis will be incomplete at best or, worse, thoroughly misleading. In my view, the same applies to abduction in natural systems. Occurring abductions are embedded in the organism's ongoing physical interactions with its environment and should be studied as such.

In recent years, especially the perceptual quality of abduction has been the subject of investigation in cognitive science, most notably in the area

1 This approach has become known under a variety of labels (e.g. enactive cognition, situated cognition, embedded cognition, distributed cognition) and has found its way in various disciplines of cognitive science such as philosophy (Clark, 1997; Juarerro, 1999), AI (Clancey, 1997), psychology (Thelen & Smith, 1994) neuroscience: Edelman, & Tononi, 1992; Damasio, 1994), and Dynamical Systems Theory (Port & van Gelder, 1995; Kelso, 1995).

of developmental psychology[2]. Several studies demonstrate the perceptual inclinations of young infants to abductions to causality and animacy. This research seems to confirm Peirce's suggestion that abduction has a strong connection to perception. I will briefly review some salient findings in order to draw some implications for the attempt to understand and model abduction in general.

Perception of causality and animacy

Scholl & Remoulet (2000) review work on how young infants abduce (a) causal relations between objects and (b) animacy (perceiving objects as being alive, having goals). Research on causality started with the work of Michotte (1946/1963), and has expanded considerably since, demonstrating the phenomenon from a very early age on and across cultures.

Experiments with 6 month olds are notoriously difficult due to the capriciousness of the subjects involved, and conclusions can be drawn only with caution. One typical experiment had 6 month olds watching simple movies of a moving object A colliding with another non-moving object B, with A then stopping and B starting to move. Such movies are known to give rise in adults to the so-called 'illusion of causality', a strong percept or conviction that A caused B's behavior (e.g. through pushing it or making it go in some other way). Basically, an explanatory hypothesis (a causal relation between A and B) is abduced, and this hypothesis helps to explain the observed phenomena (the movements of A and B). Infants, after some repetitions of this movie, show signs of habituation, i.e. they loose attention and start looking elsewhere. However, their attention is immediately captured again when a new movie is shown in which the relation between the hitting object and the object hit is inversed, and thereby their causal relationship (i.e. the old movie is played backwards). In another experimental condition, a different type of movie was shown initially, for instance displaying a time delay between the collision and the subsequent movement of the 2nd object (thereby weakening

2 I do not wish to imply that the authors discussed below unequivocally concur with the views on the central system-module distinction and the value of EEC that I am espousing here. I merely wish to analyze the results of their investigations from the perspective of EEC.

the illusion of a causal connection between A and B). In this case, the rise in attention for its reversed version was significantly lower than in the 'illusion of causality'-condition. The main investigators of these phenomena offer the following interpretation:

> "Reversal of an apparently causal event produced more recovery of attention following habituation that the reversal of a similar but apparently non-causal event (delayed reaction). (..) Hence the infant's percept of direct launching may involve more than an encoding of its spatiotemporal properties. Since the same kind of stimulus gives rise to a causal illusion in adults, it may be that the additional factor at work is the perception of a causal relationship. Leslie & Keeble (1987, p.265).

Obviously the interpretation has to be carefully phrased, as it is based on abductive reasoning itself. Yet, experiments such as these provide indications that

> "there is a rather humble perceptual mechanism operating automatically and incorrigibly upon the spatio-temporal properties of events yet producing abstract descriptions of their causal structure. It also implies that the idea of cause and effect does not originate in prolonged learning" Leslie (1988, p.186-187).

Work on animacy started with Heider & Simmel (1944). In a characteristic experiment a subject views a movie in which a simple geometric shape (e.g. a large red triangle) moves along a path that is also followed by another object (a small yellow triangle) for most of the time. At the end of the movie the small object enters a semi-closed rectangular space while the large red triangle does not follow, but stays outside, moving back and forth. In adults, such movies give rise to the perception of the objects involved as being 'animated' in the sense that goals and mental states are mentioned in a description of what is seen (e.g. the large red triangle wants to capture the small yellow triangle, but the latter wants to escape and tries to hide inside the box). Moreover, personality traits (shyness, being a bully) and emotions are attributed to the objects involved, by subjects of a wide range of cultures and for as young as 3-4 year old (Berry & Springer, 1993). Indeed, Csibra et

al. (1999) reported evidence, based on similar movie-based habituation studies, for goal attribution by infants of only 9 months old.

Conclusion

Some experimental indications have been found that abduction to causality and animacy takes place in a way not unlike perception, that is fast, automatic, mandatory and bottom up (highly stimulus-driven). Moreover, the behavioral phenomena observed during experiments are thought to be in contrast with the traditional view on abduction as a kind of high-level, central process:

> "such phenomena reflect primarily perceptual and perhaps modular processing, and at a minimum are very different (and can be dissociated) from high-level cognitive judgments of the existence of causality or animacy" (Scholl & Tremoulet, 2000, p.307).

I take these experiments and their interpretation to be in support of Peirce's suggestion that perception and abduction shade imperceptibly into one another. Moreover these results make sense from the perspective of Embodied Embedded Cognition, according to which presumed 'higher' cognitive processes are intrinsically and dynamically coupled to perception and action. Even though in the experiments discussed a proper action-component is still, and unfortunately, missing, at least the connection between perception and abduction has been revealed.

If one accepts a strong perceptual component in abduction, there may be some interesting consequences to be drawn concerning the understanding of abduction in scientific contexts. Like every other human activity, science involves doing and perceiving things. As is well known, when one gets stuck in the investigation of a particular problem, a major concern is to give oneself a different perspective, to approach the problem from a new direction. This emphasis on actively taking a different look at the problem could very well turn out to have a literal meaning, in the sense that the scientist is trying to get his or her perception going by searching for an alternative kind of self-stimulation, either through imagination or genuine vision. This way,

searching for an appropriate explanatory hypothesis may often be based on a mixture of perception and thought. One example that comes to mind is Kepler's way of finding out about the earth's orbit. He imagined himself to be on Mars, and wondered, how, from that perspective, the earth's orbit would look like. As Koestler (1959, p.329) describes it, Kepler's task

> "was to determine, more precisely than Copernicus had done, the earth's motion round the sun. For that purpose he designed a highly original method of his own. (…) It consisted, essentially, in the trick of transferring the observer's position from earth to Mars, and to compute the motions of the earth exactly as an astronomer on Mars would do it."

This type of imagining is frequent in scientific abductions (see, for instance, Shelley, 1996), and can be considered as a kind of semi-visual self-stimulation.

This perspective on the relation between perception and abduction also helps to understand an important reason underlying science's methodological constraints. As we all know, we are too susceptible to fall for the 'obviously given', the strong feeling of conviction that accompanies the recognition of a pattern in stimuli. The methodology of science with all its careful checks and double checks may have been developed not just to protect against errors of thought, but also against strongly perceptually driven abductions in the wrong directions. That is to say, it may be there to help us avoid or correct 'misperceived' explanatory hypotheses.

From a more general perspective, Peirce's suggestion and the psychological investigations discussed above may be interesting because they provide reason to reconsider the opposition between descriptive and normative aspects of human, in this case particularly abductive, reasoning. A psychological-descriptive approach investigates how, *in actual practice*, plausible hypotheses are generated and how it is determined which ones are more or less plausible. For a philosophical-normative investigation, the question is how the choice of a hypothesis can be *justified*, and whether it can be accepted as true. Whether human reasoning can ultimately be shown to be rational is of no inherent concern to psychology. It 'merely' aims at modeling our *actual*

cognitive processes. Psychology is interested in how we, in fact, reason, irrespective of how rational these processes are. Contrary to the quotes given at the beginning of this paper, it may be suggested that even though abductive thinking may be bound to follow its own ways, these ways need not be entirely 'dark'. A proper investigation of the embodied embedded aspects of abduction can help to shed light on our ability to guess wonderfully right at least some of the time (Peirce, CP 5.591, 7.219, 8.238). Abduction, though a problem, need not be such a terrible problem for cognitive science after all, once the exclusive focus on purely internal information processing is dropped.

Acknowledgements

I would like to thank Olga Pombo and Alexander Gerner for organizing an inspiring international meeting on 'Abduction and the Process of Scientific Discovery' that provided the impetus for this paper. I am grateful to Iris van Rooij and Jelle van Dijk for their useful comments on an early draft.

References

Berry, D.S., & Springer, K. (1993). Structure, motion, and preschoolers perception of social causality. *Ecological Psychology, 5*, 273-283.

Clancey, W.J. (1997). Situated *cognition: On human knowledge and computer representation*. Cambridge: Cambridge University Press.

Clark, A. (1997). *Being There: Putting brain, body and world together again*. Cambridge, Massachusetts: MIT- Press.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition*, *72*, 237-267.

Damasio, A.R. (1994*). Descartes´ error: emotion, reason, and the human brain*. New York: Avon Books.

Edelman, G.E. & Tononi, G. (2000). *A universe of consciousness: how matter becomes imagination*. New York: Basic Books.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Guha, R. V. and D. B. Lenat (1990). Cyc: a midterm report. *AI Magazine*: 32-59.

Guha, R. V. and D. B. Lenat (1993). Re: Cycling paper reviews. *Artificial Intelligence*: 149-174.

Hanson, N.R. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press

Haselager, W.F.G. (1997). *Cognitive science and folk psychology: The right frame of mind*. London: Sage.

Haselager, W.F.G., & Van Rappard, J.F.H. (1998). Connectionism, systematicity, and the frame problem. *Minds and Machines, 8* (2), 161-179.

Holland, J. H., K. J. Holyoak, Nisbett, R.E., & Thagard, P. (1986). *Induction*. Cambridge, MIT Press.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243-249.

Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge, MA: MIT Press.

Kelso, S. (1995). *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Koestler, A. (1959). *The sleepwalkers*. London: Penguin.

Leslie, A.M. (1988). The necessity of illusion: Perception and thought in infancy. In: L. Weiskrantz (Ed.). *Thought without language*. Oxford: Clarendon Press.

Leslie, A.M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265-288.

Michotte, A. (1946/1963). *The perception of causality*. New York: Basic Books.

Peirce, C. S. (1931-1958). Collected papers (CP), v. 1-8. In C. Hartshorne, P. Weiss and A. Burks (Eds.). *The collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Port, R.F. & van Gelder, T. (Eds.). *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Putnam, H. (1992). *Renewing philosophy*. Cambridge, Harvard University Press.

Pylyshyn, Z. W., Ed. (1987). *The robot's dilemma*. Norwood, ABLEX Publishing Corporation.

Reichenbach, H. (1947). *Elements of symbolic logic*. New York, The Free Press.

Scholl, B.J. & Tremoulet, P.D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 8: 299-309.

Shelley, C. (1996). Visual abductive reasoning in archeology. *Philosophy of Science, 63*, 278-301.

Singh, P. & Minsky, M. (2002). A ten year roadmap to machines with common sense *Report for the National Science Foundation*.

Singh, P. Minsky, M. & Eslick, I. (2004).An architecture for computing commonsense. *BT Technology Journal, 22*(4):201-210.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MIT-Press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thelen, E., & Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Massachusetts: MIT-Press.

Clinical rationality in Medicine: The role of abductive inference

João Sampaio
Centro de Filosofia da Universidade de Lisboa
cmhiperbarica@vizzavi.pt

According to Charles Sanders Pierce pragmatic utmost principle « Elements of every concept reach the logic through the perception gate and they exit by the utilitarian action gate. Every thing that does not disclose its passports at both gates will face attainment denied by the reason as a non authorized entity" (The collected papers-212).

This work aims at analysing the practical use of this pragmatic principle in medical practice. We bring into special prominence the outcome of clinical reasoning and therefore the diagnosis (concept) to a logic idea based on the clinical history (perception) and serves to define therapy (utilitarian action).

We conclude that within the clinical reason model coexist abductive, deductive and inductive types of inferences. However only the abduction taken by its double epistemological meaning works as "passport for the perception and utilitarian action gates".

I will present medical reasoning as a typical example, that uses, as essential strategy, the abductive inference and the logic of Charles Pierce.

According to Charles Pierce's pragmatic utmost principle, "Elements of every concept, reach the logic though through the perception gate and they exit through the utilitarian action gate. Everything that does not disclose its passport at both gates, will face attainment denied by the reason as a non authorized entity". Following this principle, we can say, that medical practice

makes use of a clinical reasoning where the concept is the diagnosis. It is the outcome of perception given by the clinical history and the physical examination, that leads to an utilitarian action, which is the treatment. Following this pathway, we want to demonstrate that the passports allow its entrance in the logic though, and are, abductive inferences.

According to the Pierce's epistemology, all the cognitive process is based in signs or generality elements.

Against other more formal semiotics, which reduce the study of the relations between the sign and the interpretant, Pierce's semiotics, is based on a sign as a first or a sense, in a position of triadic relation with a second, called object or phenomenon, producing a third step in the reaction denominated interpretant.

To Pierce, the concept of Sign overcomes the idea of simply sensitive, when it implies, a triadic relation, organised and complete, between sign, object and interpretant. When this concept of sign is followed, the mind (or semiosis, according to Pierce), becomes a process of unlimited meanings, reason why, a certain Third in a triadic relation, becomes a First, in another triadic relation.

This process of unlimited generation of meanings, allows the object included in a certain sign, goes from representative (the immediate object) to dynamic object, corresponding to the object as showed in a final and unlimited study.

Likewise, the interpretant, as a mental image created within the sign triadic relation, evolves in a similar fashion. Taken initially as immediate, or a meaning without any previous reflection, it evolves progressively becoming and reaching a dynamic meaning. At last, it acquires the status of self corrected mechanism, named final interpretant.(Table 1)

In clinical practice, the semiotic process is elaborated using the clinical history and the objective examination.(P.D. Welby - 2002)

The clinical history includes:

- Patient's identification and respective social and cultural background
- Reason why the patient attended medical support
- Actual disease history
- Personal background
- Family background
- Personal habits
- Questionnaire conducted by the physician investigating organs and systems

The objective examination includes:

- General examination
- Organs and systems topographic examination

According to Pierce's semiotic model, that process aims at relating the patient with a dynamic object, and the physician with a final interpretant. The more this process takes place, the better becomes the physician's capacity, as an interpreter, to take the ultimate step towards clinical reasoning. This process involves perception data, diagnostic hypotheses, and discussion of the same hypotheses.

As we will try to demonstrate, this step makes use of abdutive inference.

The concept of abduction has been initially described by Charles Sanders Pierce in 1903 as a "process of forming an explanatory hypothesis". And he added "it is the only logical operation which introduces any new idea".

Having started with Pierce, this concept has been broadened and has lately with Magnani in 2001 acquired a double epistemological meaning: First, as inference that generates plausible explanatory hypotheses of phenomena or observations. Secondly, as an inference giving rise to a better explanation amongst several plausible hypotheses.

In rational diagnostic, the elaboration of diagnostic hypotheses, and its differential discussion, based on the clinical history and objective examination follows a strategy where both epistemological meanings are present. The elaboration of each diagnostic hypotheses, is performed conjugating

the group of signs that lead to the acceptance of the given hypotheses, and also, the signs description that go against the acceptance of the same hypotheses.

In Pierce's conception, the signs are not pure, and are full of theories. It is exactly what happens in the diagnostic reason, where the diagnostic abduction, is brought into context (Michael Hoffman - 1997) by the usual use of a thirdness sign, called symbol. (To Pierce, symbol is a sign referring to an object, due to an association of ideas based on a convention or a theory). So, the role of the instinct turns quite reduced.

Hence, there is a double epistemological meaning to the diagnostic abduction. Namely: diagnostic hypotheses elaboration, and qualitative graduation of each of the given diagnostic hypotheses.

Once the diagnostic hypotheses are established, and a differential diagnosis is achieved, complementary exams with a logical meaning to all given hypothesis are done. For example: if the diagnostic hypothesis is anaemia, blood tests will show a low haemoglobin level. This inference is deductive, and applies for consequences

Put together, all the complementary tests results are, by the means of inductive inference (type III according to Pierce), do we, or don't we confirm, the given diagnostic hypotheses, hence allowing the establishment of a final diagnosis.

There is a correspondence between the diagnoses and normative therapy based in the consensus, the so called Evidence Based Medicine. However, the singularity of disease phenomenon, justifies a non normative therapy, which is made, by using in the context, the so called Pierce's human being as sign. Following abduction, we will reach plausible therapeutic hypotheses and the best of them, will be disclosed to the patient.(Table II)

We conclude that within the clinical reason model, coexist, abductive, deductive and inductive types of inferences. However, only the abduction taken by its double epistemological meaning, (inference generator of ex-

planatory hypothesis for phenomenon or observations, and inference generator of the best explanation within multiple plausible hypothesis), works as passport for the perception and utilitarian action gates.

References

1. Hoffman, M (1997): *Is there a "Logic" of Abduction?*. Proceedings of the 6th Congress of the IASS-AIS, I International Association for Semiotic Studies in Guadalacara, México, July, 13-18, 1997.
2. Magnani, L. (2001): *Abduction, reason and science. Processes of discovery and explanation*. New York: Kluwer Academic/Plenum Publishers.
3. Peirce, C.S., *Collected Papers*, Harvard University Press, Cambridge, MA, 1931-1958.
4. Welsby, P.D. (2002): *Clinical History Taking and Examination*. Harcourt Publishers, Edimburgh.

Scientific discovery and social symptomatology: abduction, medical semeiotics and semioethics

Susan Petrilli
Dipartimento di Pratiche linguistiche e Analisi di Testi
Università degli Studi di Bari
s.petrilli@lingue.uniba.it

Premise

In this paper global semiotics is developed in the direction of what we propose to call *semioethics*.

As a unique semiotic animal, that is to say, the only animal capable of reflection upon signs and communication, the human being has a singular responsibility towards life (which converges with signs and communication), which also means the quality of life. More than *limited responsibility*, the type of responsibility involved is *unlimited responsibility*, *responsibility without alibis*, *absolute responsibility*.

The capacity for abduction as the condition for critical and dialogic totalization implies that the ability to grasp the *reason* of things cannot be separated from the capacity for *reasonableness*. The issue at stake may be stated in the following terms: given the risks inherent in social reproduction today for semiosis and for life*, human beings must at their very earliest transform from rational animals into reasonable animals*.

It is significant that Peirce too should have turned his attention specifically to the normative sciences in the final phase of his reseach. He linked logic to ethics and to aesthetics: while logic is the normative science concerned with self-controlled thought, ethics is the normative science that focuses on self-controlled conduct, and aesthetics the normative science devoted to ascertaining the end most worthy of our espousal. In this context,

Peirce took up the question of the ultimate good, *summum bonum,* or ultimate value which he refused to identify with either individual pleasure (hedonism) or with a societal good such as the greatest happiness for the greatest number of human beings (English utilitarianism). Rather, he insisted that the *summum bonum* could only be defined in relation to the 'evolutionary process', that is, to a semiosic process of growth. Specifically, he identified the highest good in the continuous 'growth of concrete reasonableness'.

Our responsibilities towards life in the global communication-production phase of development in late capitalist society are enormous, indeed unbounded. This is also true in the sense that when we speak of life the implication is not only human life, but all of life throughout the whole planetary ecosystem, from which of course human life cannot be separated. As the study of signs semiotics cannot evade this issue.

Originally, semiotics was understood as 'semeiotics' (a branch of the medical sciences) and was focused on symptoms. Nowadays the ancient vocation of semiotics as it was originally practiced for the 'care of life' must be recovered and reorganized in what we propose to call 'semioethic ' terms. This issue is particularly urgent in the present age in the face of growing interference in communication between the historical-social sphere and the biological sphere, between the cultural sphere and the natural sphere, between the semiosphere and the biosphere.

Semioethics is the result of two thrusts: one is *biosemiotics*, the other *bioethics*.

From this point of view semiotics is *listening*, it is turned to listening, this time not in the sense of the general theory of signs subtending semiotics, but in a medical sense, semiotics must be listening in the sense of medical semeiotics or symptomatology. Semiotics must listen to the symptoms of today's globalized world and identify the different expressions of unease and disease – in social relations, in international relations, in the life of single individuals, in the environment, in life generally over the entire planet.

The aim is to make a diagnosis, a prognosis and to indicate possible cures to the end of creating a future for globalization, for the health of semiosis in opposition to a globalized world tending towards its own destruction.

My paper starts from Thomas A. Sebeok (1920-2001) and his conception of semiotics integrated with Charles Sanders Peirce's pragmatic concept of 'abduction', disregarded by Sebeok in his own original reformulation of Peircean theory. Sebeok was not interested in logic just as he was not interested in critiquing today's social system. Sebeok denominated his particular approach to semiotics as 'global semiotics'. Taking global semiotics as our starting point, we propose to develop it in the direction of *semioethics*, which presupposes Sebeok's interpretation of ancient medical semeiotics or symptomatology as an initial phase in the history of semiotics.

Abduction, icon, and agapasm from a synechist perspective

Abduction is the name of a given type of argumentation, of progression from one interpretant to another. Abduction is foreseen by logic, but (especially in its more risky expressions) supercedes the logic of identity and develops through argumentative procedures that are risky, that is, eccentric, creative, or inventive. By contrast with induction and deduction, in abduction the relation between interpreted sign and interpretant sign is regulated by similarity, attraction and reciprocal autonomy. Abduction is grounded in the logic of otherness, substantial dialogism, and creativity. It proceeds through relations of fortuitous attraction among signs and is dominated by iconicity.

We have made the claim that abductive argumentative procedure is risky; in other words, it advances through arguments that are tentative and hypothetical, leaving a minimal margin to convention and to mechanical necessity. To the extent that it transcends the logic of identity and equal exchange among parts, abduction belongs to the side of excess, exile, *dépense*, giving without a profit, the gift beyond exchange, desire. It proceeds more or less according to the 'interesting,' and is articulated in the dialogic and disinterested relation among signs – a relation regulated by the law of creative love, so that abduction is also an argumentative procedure of the agapastic type.

In Charles S. Peirce's philosophical system (see, for example, his writings collected in the volume *Chance, Love and Logic*, edited by M. R. Cohen, 1923), 'chance', 'love' and 'necessity' indicate three modes of development. The

problematic of love, or what Peirce *agapasm*, is interesting for its connection to abductive inference.

In his renown paper 'On a New List of Categories' (*CP* 1.545-559), Peirce elaborates his doctrine of categories with the triad firstness, secondness, and thirdness, which are always co-present, interdependent and irreducible. The doctrine of categories constitutes the foundation of his ontology and cosmology, therefore it is connected to his ontological-cosmological triad with the distinction with agapasm, anancasm, tychasm, to his typology of inferential logic with the triad abduction, induction, and deduction, in addition to his sign typology, in particular the triad icon, index, and symbol.

The connection between agapasm and abduction offers a platform to discuss Victoria Welby's (1837-1912) work in relation to Peirce. In a reconstruction of the history of semiotics, Welby is a name to remember along with others like Mikhail M. Bakhtin and Emmanuel Levinas for a better understanding of the sign in theoretical terms.

Understood as development through the forces of affinity and sympathy and referring to the Peircean triad, icon, index and symbol, we may claim that agapasm is strongly iconic by virtue of the force of attraction, that is, the relation of similarity or affinity among interpretants. In agapastic evolution, chance (tychasm) and necessity (anancasm) are operative, but the forces of attraction, affinity, freedom, and fortuitousness dominate in the relation among interpretants forming the continuous (synechetic) flow of infinite semiosis, together with iconicity over indexicality and symbolicity.

The concept of continuity or synechism involves regularity. As emerges from her own philosophy of the signifying processes permeating the entire universe, Welby too believed that development is articulated in a structure, and that continuity presupposes relational logic grounded in otherness. The logic of otherness is a dia-logic – that is, logic that recovers the dimension of *dialogicality*, as understood by Bakhtin. In other words, following both Peirce and Bakhtin, dialogicality is considered as a modality of semiosis, which may or may not involve verbal signs and may or may not take the form of dialogue. Dialogicality thus described is determined by the degree of opening towards otherness. And agapastic evolution is achieved through the law of

love; creative and altruistic love, love founded on the logic of otherness, as would say authors like Welby, Bakhtin and Levinas.

In tychastic development – connected to symbolicity in semiotic terms and to induction in argumentative terms, new interpretive routes with unpredictable outcomes, some of which are fixed in 'habits,' are determined by chance. In anancastic development – connected to indexicality and deduction, new interpretive routes are determined by necessity, by causes that are both internal (the logical development of ideas, of interpretants that have already been accepted) and exteznal (circumstances) with respect to consciousness, without the least possibility of prediction concerning eventual results. Instead, in agapastic development, deferral from one interpretant to the next is iconic and abductive. Development is neither regulated by chance, nor by blind necessity, but, as Peirce says, 'by an immediate attraction for the idea itself, whose nature is divined before the mind possesses it, by the power of sympathy, that is, by virtue of the continuity of mind' (*CP* 6.307). As an example of agapasm, the evolution of thought according to the law of creative love, Peirce cites the *divination* of genius, the mind affected by the idea before it is comprehended or possessed, by virtue of the force of attraction that the idea exercises upon him in the context of relational continuity among signs in the great semiosic network of the universe or semiosphere.

Paradoxically, in tychastic development chance generates order. In other words, the fortuitous result engenders the law, and the law in turn finds an apparently contradictory explanation in terms of the action of chance. This is the principle that informs the work of Charles Darwin on the origin of the species and natural selection. In Peirce's opinion, one of the reasons why Darwin won so much favor was that the values informing his research – represented by the principle of the survival of the fittest – measured up to dominant values of the time, which, in the final analysis, were founded on the assertion of the logic of identity and could be summed up in the word 'greed'.

Logic understood in the strict sense as necessary cause is connected with anancastic development. The limitation to this kind of development is the conviction that only one kind of logical development is possible. By main-

taining that the conclusion derived from the premises could not be anything different, all other argumentative modes, and consequently the possibility of free choice, are excluded (see *CP* 6.313). However, while the mechanisms of constriction, contingency, and mechanical necessity effectively dominate over the relation between interpreted and interpretant signs in anancastic inferential procedure, this does not exclude the effect of other interpretive modalities, active even when the anancastic procedure prevails. In semiotic terms, in the case of anancasm the relation between the intepreted and interpretant is predominantly indexical; in argumentative terms, it is deductive. The relation between the conclusion and its premises is regulated by reciprocal constriction and operates at low degrees of otherness and dialogism.

The goal of agapastic development is development itself (of the cosmos, language, thought, the subject), continuity of semiosis, of the development of an idea. According to Peirce, in a universe regulated by the principle of continuity and relational logic, where no single fact, datum, idea or individual exists in isolation, creative evolution (beaten out at the rhythm of hypotheses, discoveries, and qualitative leaps) is achieved through the combined effect of agapasm and synechism. Therefore, shifting our attention to the problem of subjectivity, the self as a communicating entity, far from being solitary in essence, has its roots in *agape*. By virtue of the continuity of thought, and therefore of relational logic, the main force ruling over the deferral among signs in an evolutionary perspective is that of *agapic or sympathetic comprehension and recognition*. And the simultaneous and independent occurrence of a genial idea to a number of individuals not endowed with any particular powers – a consequence of belonging to the same great semiosphere – might be considered as a demonstration of this (see *CP* 6.315-316). As a sign, in Peirce's view, the self too develops according to the laws of inference (see *CP* 5.313).

Medical semeiotics and semioethics: from Sebeok's biosemiotics and beyond

Let us now consider Sebeok's 'global semiotics'. According to Sebeok, semiotics is not only *anthroposemiotics* but also *zoosemiotics*, *phytosemiotics*, *mycosemiotics*, *microsemiotics*, *endosemiotics, machinesemiotics*, *environmental semiotics*.

Sebeok's interests cover a broad range of territories ranging from the natural sciences to the human sciences. Consequently, he deals with theoretical issues and their applications from as many angles as the disciplines called in question: linguistics, cultural anthropology, psychology, artificial intelligence, zoology, ethology, biology, medicine, robotics, mathematics, philosophy, literature, narratology, and so forth. Even though the initial impression might be of a rather erratic mode of proceeding as he experiments various perspectives and embarks upon different research ventures, in reality Sebeok's expansive and seemingly distant interests find a focus in his 'doctrine of signs' and in the fundamental conviction subtending his general method of enquiry that the universe is perfused with signs, indeed, as Peirce hazards, may be composed exclusively of signs.

As a fact of signification the entire universe enters Sebeok's 'Global Semiotics'. Semiotics is the place where the 'life sciences' and the 'sign sciences' converge, therefore the place where consciousness is reached of the fact that the human being is a sign in a universe of signs.

Sebeok has propounded a wide-ranging vision of the study of signs that coincides with the study of the evolution of life. After Sebeok's work both the conception of the field of semiotics (the sign science) and its history are changed noticeably. Thanks to him semiotics at the beginning of the new millennium presents a far more enlarged view than that of the first half of the 1960s. Sebeok has extended the boundaries of traditional semiotics or more correctly semiology – which is restrictively based upon the verbal paradigm and is vitiated by the *pars pro toto* error. He tagged this conception of semiotics the 'minor tradition' and promoted, instead, what he called the 'major tradition', as represented by Locke and Peirce and early studies on signs and symptoms by Hippocrates and Galen. Semiotics, therefore, is at

once recent if considered from the viewpoint of the determination of its status and awareness of its wide-ranging possible applications, and ancient if its roots are traced back at least, following Sebeok (1979) to the theory and practice of Hippocrates and Galen.

By virtue of this 'global' or 'holistic' approach, we can immediately associate Sebeok's inquiry into the 'life of signs' with his concern for the 'signs of life'. In Sebeok's view, *semiosis and life converge*.

The boundaries of the history of semiosis have also been extended as a result of the links forged between semiosis and the biological sciences, a consequence of including not only communication but also signification and symptomatization in our conception of semiosis or sign process or sign situation.

Moreover, the boundaries of the history of semiotics are further expanded by awareness of the interdisciplinary character of semiotics, including its relation to the natural and biological sciences.

Semiotics as a field of study is far older than is usually presented. Its origins are in fact ancient, notwithstanding superficial descriptions in a phonocentric and glottocentric key. Semiotics grew out of symptomatology, or medical semeiotics, which Galen considered one of the principal branches of medicine. In this sense, Hippocrates and Galen, with their early studies on signs and symptoms, were among the very first semioticians.

If semiotics is concerned with life over the whole planet (given that life and semiosis converge), and if the original motivation for the study of signs is the 'health' of semiosis, it may well follow that an important task for semiotics today is to care for life in all its diversity – a task that is especially urgent in the era of globalization. Semiotics is capable of understanding the entire semiosic universe and of discussing the various forms of separatism, technicalism, and overspecialization that afflict it. All of this should result in awareness of ethical responsibility.

*Semioethic*s is the term we have proposed for this particular unbounded trend in semiotics, as announced in my book with Augusto Ponzio entitled *Semiotics Unbounded*. This title is intended to underline the totalizing and yet

open character of semiotics, with what we may call its 'detotalizing method'.

Semiotics is an extremely wide-ranging field that crosses into many disciplines. Precisely because of this, and given that we must keep account of progress in the sciences – 'human', 'natural' and 'logico-formal' sciences –, semiotics must be ready to renew itself and to interrogate the very methods and categories it employs. Semiotics is unbounded, and so is the object of its studies – the sign network. This leads to the subtitle of our book: *Interpretive routes through the open network of signs*.

In a paper of 1989, 'Semiosis and Semiotics: What lies in Their Future?,' included in *A Sign Is Just a Sign* (pp. 97-99), Sebeok introduced a second meaning of the term 'semiotics' beyond indicating the general science of signs. This new meaning refers to *the specificity of human semiosis*, the human capacity for metasemiosis, and is of vital importance for a *transcendental founding of semiotics* as a doctrine of signs. Says Sebeok:

> Semiotics is an exclusively human style of inquiry, consisting of the contemplation – whether informally or in formalised fashion – of semiosis. This search will, it is safe to predict, continue at least as long as our genus survives, much as it has existed, for about three million years, in the successive expressions of Homo, variously labelled — reflecting, among other attributes, a growth in brain capacity with concomitant cognitive abilities — *habilis, erectus, sapiens, neanderthalensis,* and now *s. sapiens*. Semiotics, in other words, simply points to the universal propensity of the human mind for reverie focused specularly inward upon its own long-term cognitive strategy and daily manoeuvrings. Locke designated this quest as a search for 'humane understanding'; Peirce, as 'the play of musement' (ivi: 97).

According to this second meaning, 'semiotics' is the human species-specific capacity for *metasemiosis*. In other words, human semiosis is capable of reflecting on signs, therefore not only of interpretation undistinguished from direct response to signs, but also of interpretation understood as reflection on signs, as suspension of response and deliberation.

Beyond Aristotle's who at the beginning of his *Metaphysics* claimed that man tends by nature to knowledge, we can now claim that man tends by

nature to semiotics. Human semiosis, anthroposemiosis, also presents itself as *semiotics* in the doube sense just described and as such is free to venture across the entire semiosic universe. All the same, however, we must guard against the pars pro toto error (made in the history of ideas), that is, the oversimplification of reducing semiosis in general to anthroposemiosis.

In another paper included in *A Sign is Just a Sign*, 'The evolution of semiosis' (pp. 83-96, now in Posner et al. 1997-1998), Sebeok explains the correspondences that exist between the various branches of semiotics and different types of semioses, from the world of micro-organisms to the Superkingdoms and the human world. Human semiosis is characterized as semiotics thanks to a modeling device specific to humans, called 'language' (it is virtually certain that *Homo habilis* was endowed with language, but not speech).

In another important paper included in *A Sign Is a Just a Sign*, entitled 'In What Sense is Language a 'Primary Modeling system'?' (now also in *Signs*, 1994), Sebeok describes language as a 'modeling device'. Every species is endowed with a model that produces its own world, and 'language' is the name of the model belonging to human beings. The human modeling device, or language, is completely different from the modeling devices of other life forms. Its distinctive feature is what the linguists call 'syntax'. Thanks to syntax hominids do not just experience one 'reality', one world, but can also frame an indefinite number of possible worlds. This capacity is unique to human beings. Thanks to syntax human language is like Lego building blocks, it can reassemble a limited number of construction pieces in an infinite number of different ways. As a modeling device language can produce an indefinite number of models; in other words, the same pieces can be taken apart and put together to construct an infinite number of different models. Thanks to language not only do human animals produce worlds as do other species, but, as says Leibniz, human beings can also produce an infinite number of possible worlds. This leads to the 'play of musement', a human capacity which Sebeok considers particularly important for scientific research and all forms of investigation as well as for fiction and all forms of artistic creation.

The exquisitely human propensity for musement implies the ability to carry out such operations as predicting the future or 'traveling' through the past, the ability, that is, to construct, deconstruct and reconstruct reality, inventing new worlds and interpretive models. (The happy expression 'the play of musement' is used by Sebeok, interpreter of Peirce, as the title of his book of 1981).

Speech appeared with *Homo sapiens* as an adaptation, like language, but *for the sake of communication*, and much later than language. Consequently, language too ended up becoming a communication device; and speech developed out of language as a derivative *exaptation* (a term proposed by Gould and Vrba 1982). Exapted for communication, first in the form of speech and later of script, language enabled human beings to enhance the nonverbal capacity with which they were already endowed. On the other hand, speech was *exapted* for modeling and eventually functioned as a *secondary modeling system*. In addition to increasing the communication capacity, speech also increased the capacity for innovation and 'play of musement'. Such aspects as the plurality of languages and 'linguistic creativity' (Chomsky) testify to the capacity of language understood as a primary modeling device, for producing numerous possible worlds.

Sebeok uses the concept of modeling as proposed by the so-called Moscow-Tartu school (A. A. Zaliznjak, V. V. Ivanov, and V. N. Toporov. Ju. M. Lotman) where it is used to denote natural language ('primary modeling system') and other human cultural systems ('secondary modeling systems'). However, differently to the Moscow-Tartu school, Sebeok goes further to extend the concept of modeling beyond the domain of anthroposemiosis. With reference to the biologist J. von Uexküll and his concept of *Umwelt*, Sebeok's interpretation of model may be translated as an 'outside world model'. On the basis of research in biosemiotics, the modeling capacity is observable in all life forms (see Sebeok 1991b: 49-58; 1994b: 117-127). *The Forms of Meaning. Modeling Systems Theory and Semiotic Analysis*, co-authored by Sebeok with Marcel Danesi, published in 2000, further develops the fundamental notion of 'model'.

The study of modeling behavior in and across all life forms requires a methodological framework developed in the field of biosemiotics, what Sebeok in his research on the interface between semiotics and biology called 'modeling systems theory'. Modeling systems theory studies semiotic phenomena as modeling processes (see Sebeok and Danesi 2000: 1-43).

In the light of semiotics conceived in terms of modeling systems theory, semiosis – which a vital characteristic of all life forms – may be defined as 'the capacity of a species to produce and comprehend the specific types of models it requires for processing and codifying perceptual input in its own way' (Sebeok and Danesi 2000: 5).

The applied study of modeling systems theory is called *systems analysis*, which distinguishes between primary, secondary and tertiary modeling systems.

The primary modeling system is the innate capacity for simulative modeling, in other words, it is a system that allows organisms to simulate something in species-specific ways (see Sebeok and Danesi 2000: 44-45). The primary modeling system specific to *Homo* is also called 'language'. The secondary modeling system is the system that subtends both indicational and extensional modeling processes. The nonverbal form of indicational modeling has been documented in various species, whereas extensional modeling is a uniquely human capacity given that it presupposes *language* (human primary modeling system) as distinguished from *speech* (human secondary modeling system) (see Sebeok and Danesi 2000: 82-85). The tertiary modeling system subtends highly abstract, symbol-based modeling processes. Tertiary modeling systems are human cultural systems (see Sebeok and Danesi 2000: 120-129).

Syntactics, deconstruction and reconstruction, production of many possible worlds, semiotics, with the ensuing capacity for evaluation, responsibility, inventiveness, planning, criticism, are all prerogatives of language. With language the being of communication finds its *otherwise*. Insofar as man is endowed with language, insofar as he is a semiotic animal, human behaviour cannot be circumscribed to communication, being, ontology. From this perspective man reveals his capacity for *otherness*. He can present himself as

other and propose other possibilities beyond the *alternatives* forseen by 'being' as the latter emerges in communication society.

Proceeding beyond Sebeok we may claim that the human being as a unique semiotic animal, that is, the only animal capable of reflection upon signs and communication, has a singular responsibility towards life (which is made of signs and communication) – which also means the quality of life. More than *limited responsibility*, the type of responsibility involved is *unlimited responsibility*, as understood by Levinas, *responsibility without alibis*, *absolute responsibility*.

Logic, ethics, and social symptomatology

The human capacity for abduction, that it, for hypothesis, is the condition for critical and dialogic totalization. It implies that the ability to grasp the *reason* of things cannot be separated from the capacity for *reasonableness*. The issue at stake may be stated in the following terms: given the risks inherent in social reproduction today for semiosis and for life*, human beings must at their very earliest transform from rational animals into reasonable animals*.

Referring to the problem of subjectivity in the framework of his pragmaticism, therefore to self considered as a set of actions, practices, habits, Peirce identified 'power' as opposed to 'force' as one of its fundamental characteristics. Power is not 'brute force' but the 'creative power of reasonableness,' which, by virtue of its agapistic orientation rules over all other forms of power (see *CP* 5.520). We could say that power, that is, the ideal of reasonableness, is the capacity to respond to the attraction exerted upon self by the other; power and reasonableness are related to the capacity for response to the other and the modality of such response is dialogue.

It is significant that Peirce turned his attention specifically to the normative sciences in the final phase of his research. He linked logic to ethics and to esthetics: while logic is the normative science concerned with self-controlled thought, ethics is the normative science that focuses on self-controlled conduct, and esthetics the normative science devoted to ascertaining the end most worthy of our espousal. In this context, Peirce took up the question of the ultimate good, *summum bonum,* or ultimate value which he

refused to identify with either individual pleasure (hedonism) or with a societal good such as the greatest happiness for the greatest number of human beings (English utilitarianism). Instead, he insisted that the *summum bonum* could only be defined in relation to the 'evolutionary process', that is, to a semiosic process of growth. Specifically, he identified the highest good in the continuous 'growth of concrete reasonableness'.

Our responsibilities towards life in the global communication-production phase of development in today's society are enormous, unbounded. Indeed, when we speak of life the implication is not only human life, but all life throughout the whole planetary ecosystem, from which human life cannot be separated. As the study of signs, semiotics cannot evade this issue.

Originally, semiotics was understood as 'semeiotics' (a branch of the medical sciences) and was focused on symptoms. Nowadays the ancient vocation of semiotics as it was originally practiced for the 'care of life' must be recovered and reorganized in what we propose to call 'semioethic' terms. This issue is particularly urgent in the present age in the face of growing interference in communication between the historical-social sphere and the biological sphere, between the cultural sphere and the natural sphere, between the semiosphere and the biosphere.

Proceeding beyond Sebeok's perspective, we are not simply alluding to the capacity for being otherwise with respect to being, but to the capacity that is specific to humans for being *otherwise than being*, that is, *otherwise than today's being as shaped in communication*, or so-called globalisation. The capacity for otherwise than being in fact subtends all possibilities of being otherwise. This capacity is characteristic of the semiotic animal and consists of the capacity to transcend being and today's communication world. This capacity renders the semiotic animal completely responsible not only for social reproduction, but also for life over the whole planet, the two things of course being inseparable. The capacity for *otherwise than being* denies the semiotic animal all possible alibis offered by interpretation, response, action, choices and standpoints whose sense and scope are limited to social communication in the world as it is, to alternatives made available by the world as it is. Instead,

the semiotic animal is endowed with, indeed characterized by a capacity for *otherness*.

Semioethics is the result of two thrusts: one is *biosemiotics*, the other *bioethics*. Insofar as it is connected with medical semeiotics, semioethics is *listening*, is oriented to listening, not only in the sense of the general theory of signs subtending semiotics, but in a medical sense; semiotics must be listening in the sense of medical semeiotics or symptomatology, semiotics must listen to the symptoms of today's globalized world and identify the different expressions of unease and disease – in social relations, in international relations, in the life of single individuals, in the environment, in life generally over the entire planet.

On the basis of abductive inferential processes the aim is to make a diagnosis, a prognosis and to indicate possible therapies for the future of globalization and the health of semiosis, by contrast with a globalized world tending towards its own destruction.

The semiotician today must be ready to interpret the symptoms of semiosis and its malfunctioning as produced by globalization in today's global communication-production society.

What we propose to call 'semioethics' must take the current phase in historico-social development as its starting point and proceed to analyze today's society rigorously and critically, today's communication-production social structures, the communication-production relationships forming the presentday world.

If semiotics is to meet its commitment to the 'health of semiosis' and to cultivate its capacity for understanding the entire semiosic universe, it must continuously refine its auditory and critical functions, that is, its capacity for listening and critique. To accomplish such tasks we believe that the trichotomy that distinguishes between (1) cognitive semiotics, (2) global semiotics, and (3) semioethics is no less than decisive not only in theoretical terms but also for reasons of a therapeutic order.

Symptomatology of global communication today from a semioethic perspective

If we consider the contribution made by global semiotics to semioethics in relation to present day global communication, sign practitioners are faced with an enormous responsibility, that of evidencing the limits of today's communication-production society. Semiotics must now accept the responsibility of denouncing incongruencies in the global system with the same energy, instruments and social possibilities produced by the global communication-production system itself – it is time to denounce the dangers involved in this system for life over the entire planet.

The current phase in the development of today's production system is 'global communication'. This expression may be understood in at least two senses: that these days communication is characterized by its *planetary extension*, and that it is *realistic in the sense that it accommodates the world as it is*. Globalization implies the *omnipresence* of communication in production and characterizes the entire productive cycle: not only is globalization present at the level of the market, of exchange, as in earlier phases in socio-economic development, but also at the level of production and consumption. Globalization is tantamount to heavy interference on the part of communication-production not only in human life, but life in general over the whole planet.

For an understanding of world-wide global communication-production we need a view that is just as global. The special sciences taken separately are unable to provide this. On the contrary, a global view is offered by the general science of signs or *semiotics* as it is taking shape today on the international scene thanks to the approach fostered by Sebeok and his ongoing work for further development.

A full understanding of the current phase in global communication implies a full understanding of the risks involved by global communication, including the risk that *communication itself may come to an end*. This risk, however, is not simply the risk of 'incommunicability', theorized and represented in film and literature, a subjective-individualistic malady ensuing from

the transition to communication in its current forms (and which cannot be separated from production). Instead, when we speak of the 'risk that communication may come to an end', we are referring above all to identification between communication and life, and therefore to the risk that life may come to an end on the planet, considering the enormous potential for destruction in today's society by contrast with all other earlier phases in the development of today's dominant social system.

Therefore, the expression *global* communication-production does not only refer to the expansion of the means of communication and of the market at a world-wide level, but also the fact that the whole of human life is englobed into the communication-production system: whether in the form of development, well-being and consumerism or of underdevelopment, poverty and the impossibility to survive; health or sickness; normality or deviation; integration or emargination; employment or unemployment; transfer functional to the work-force characterizing emigration or transfer of peoples in their denied request of hospitality, characteristic of migration; whether in the form of the traffick and use of legal or illegal merchandise, from drugs to human organs, to 'non-conventional' weapons. Indeed, this process of englobement is not limited to human life alone. All life over the entire planet is now irremediably involved (even compromised and put at risk) in the communication-production system.

Reflection on problems relevant to semioethics today in the context they in fact belong to, the context of globalization, requires an approach that is just as global: an approach that is not limited to considering partial and sectorial aspects of the communication-production system according to internal perspectives functional to the system itself, therefore an approach that on an empirical level is not limited to psychological subjects, subjects reduced to parametres imposed by the social sciences, subjects measurable in terms of statistics. Global communication-production calls for a methodological and theoretical perspective that is as global as the phenomenon under observation, in other words, a perspective that understands the logic of global communication-production, and therefore can proceed to a *critique* of the system it subtends.

Social reproduction in the global communication-production system is destructive. Reproduction of the *productive cycle* itself is destructive. It destroys: a) machines, which are continuously substituted with new machines – not because of wear, but for reasons connected with competetivity; b) jobs, making way for automation which leads to an increase in unemployment; c) products on the market where new forms of consumerism are elicited, completely ruled by the logic of reproducing the productive cycle; d) already existing products which once purchased would otherwise exhaust the demand, and which in any case are designed to become outdated and obsolete almost immediately, as new and similar products are continuously introduced on the market; e) commodities and markets which are no longer able to resist competition in the global communication-production system.

Communication-production destroys natural environments and life forms. It also destroys different economies and cultural differences eliminated by the processes of homologation operated by market logic: nowadays, not only are habits of behaviour and needs but even desires and the imaginary homologated and rendered identical (though the possibility of satisfying such needs is never identical). Communication-production also destroys traditions and cultural patrimonies that contrast with or obstacle or are simply useless, non functional to the logic of development, productivity and competition. Communication-production destroys productive forces that escape the limits of forms of production that penalize intelligence, inventiveness and creativity by subjecting them to 'the reason of the market' (and which of course cannot be avoided in the current phase of necessary investment in 'human resources'). The destructive character of today's production system is also manifest in the fact that it produces growing areas of underdevelopment as *the very condition for development,* areas of human exploitation and misery to the point of nonsurvival. This logic is behind the expanding phenomenon of *migration* which so-called 'developed' countries are no longer able to contain owing to objective internal space limits – far more serious than in earlier phases in the development of social organizations.

The principle of exploiting other people's work is destructive, work costs less the more it produces profit: with the help of global communica-

tion developed countries turn to low cost work in underdeveloped countries ('stay where you are and we will bring you work'). The disgrace of the communication-production world is manifest in the spreading exploitation of child labour that is heavy and even dangerous (much needs to be said and done about children as today's victims of underdevelopment – in misery, sickness, war, on the streets, in the work-force, on the market).

The destructive character of world-wide communication-production is evidenced by war which is always a scandal. Global communication-production is the communication-production of war. War requires ever new markets for the communication-production of conventional and unconventional weapons. War also demands to be acknowledged as just and necessary, a necessary means of defence against the menacing 'other', of achieving respect for the rights of one's 'own identity', 'one's own difference'. The truth is that identities and differences are not at all threatened or destroyed by the 'other', but rather by today's social system which while encouraging and promoting identity and difference renders them *fictitious* and *phantasmal*. And this is what pushes us to cling to such values so passionately, being a logic which fits the communication-production of war to perfection.

With the spread of 'bio-power' (Foucault) and the controlled insertion of bodies into the production apparatus, world communication goes hand in hand with the spread of the concept of the individual as a separate and self-sufficient entity. The body is perceived as an isolated biological entity, as belonging to the individual, as a part of the individual's sphere of belonging. This has led to the quasi total extinction of cultural practices and worldviews based on intercorporeity, interdependency, exposition and openning of the body (what remains is the expression of a generalized tendency to museumification; mummified remains studied by folklore analysts, archeological remains preserved in ethnological museums and in the histories of national literatures).

Technologies of separation are applied to human bodies, interests, to individual and collective subjects and are functional to production and to making production and consumption converge, which is a characteristic of presentday production forms. Instead, thanks to its ontological perspective

global semiotics (or semiotics of life) can if nothing else oppose a whole series of signs showing how each instant of individual life is interrelated, even compromised with all other life forms over the planet.

To acknowledge such interrelatedness, such compromise involves a form of responsibility which transcends positive rights and limited responsibilities, responsibilities with alibis. And to acknowledge interrelatedness is ever more urgent the more the reasons of production and global communication functional to it impose ecological conditions that obstacle and distort communication between our bodies and the environment.

References

Caputo, Cosimo; Petrilli, Susan and Augusto Ponzio (2006). *Tesi per il futuro anteriore della semiotica*. Milan: Mimesis.

Deely, John (1998). 'Thomas A. Sebeok', in Paul Bouissac (ed.), *Encyclopedia of Semiotics*,
Oxford University Press, Oxford.

Deely, John; Petrilli, Susan and Augusto Ponzio (2005). *The Semiotic Animal*. Ottawa, Toronto, New York: Legas.

Gould, Stephen J.; Vrba, Elizabeth (1982). 'Exaptation', *Palebiology* 8, I, pp. 14-15.

Lévi-Strauss, Claude (1986). 'Avant-Propos', in P. Bouissac, M. Herzfeld, R. Posner (ed.), *Iconicity. Essays on Nature and Culture. Festschift for Thomas A. Sebeok on the his 65th birthday*, Stauffenburg, Tübingen1986.

Petrilli, Susan (1986). 'On the Materiality of Signs', *Semiotica* 62, 3/4, pp. 223-245.

—(introd., trad., e cura di) (1987a). *Per Ferruccio Rossi-Landi. Il Protagora* XXVII, 11/12.

—(1987b). 'Il contributo di Rossi-Landi allo studio di Charles Morris', *Per Ferruccio Rossi-Landi. Il Protagora* XXVII, 11/12, pp. 105-114.

—(1990). 'On the Materiality of Signs', in A. Ponzio 1990a, pp. 365-401.

—(1991). 'From Peirce (via Morris and Jakobson) to Sebeok: Interview with Thomas A. Sebeok', in T. A. Sebeok 1991a, pp. 95-105.

—(ed. and intro.) (1992). *The Correspondence Between Morris and Rossi-Land. Semiotica,* Special Issue 88, 1/2.

—(1994). 'Signs and Values in Linguistic Work', in J. Bernard *et alii* (ed.), *Reading su Ferruccio Rossi-Land. Semiosi come pratica sociale*, pp. 135-159, Edizioni Scientifiche Italiane, Naples 1994.

—(1995a). *Materia segnica e interpretazione*, Milella, Lecce.

—(1995b). 'For a Global Approach to Semiosis', *Cruzeiro Semiotico*, Sept., pp. 27-36.

—(1995c). 'La Iconicidad en la 'Doctrina de los signos' de Thomas A. Sebeok', *Cruzeiro Semiotica,* Sept., pp. 303-313.

—(1996). 'Communication, Mass Media and Critique of Ideology: Ferruccio Rossi-Land's Philosophy of Language', *Socio-Semiotics. S-European Journal for Semiotic Studies* 7, 1-2, pp. 165-188.

—(1998). *Teoria dei segni e del linguaggio*, Graphis, Bari 2001[2].

—(1999). 'Charles Morris's Biosemiotics', *Semiotica* 127-1/4, pp. 67-102.

—(2003). Bodies and signs. For a typology of semiosic materiality, in S. Petrilli e P. Calefato (eds.), 2003a, pp. 181-200.

—(2005). *Percorsi della semiotica*, Graphis, Bari.

Petrilli, Susan; Ponzio, Augusto (1994). 'Exchange in Alice's World', in C. Marello and R. Fordyce (eds.), *Semiotics and Linguistics in Alice's Worlds*, Walter de Gruyter, Berlin, 74-78.

—(1998). 'Signs of Research on Signs. Semiotische Berichte', *Österreichschen Gesellschaft für Semiotik*. Special Issue Jg. 22, 3/4.

—(1999). *Fuori campo. II segni del corpo tra rappresentazione ed eccedenza*, Mimesis, Milan.

—(2000). Il sentire della comunicazione globale, Meltemi, Rome.

—(2001). S*ebeok and the Signs of Life*, Icon Books, London.

—(2002a). *I segni e la vita. La semiotica globale di Thomas A. Sebeok*, Spirali, Milan.

—(2002b). 'Sign Vehicles for Semiotic Travels. Two New Handbooks', *Semiotica* 141-1/4, pp. 203-350.

—2003). *Semioetica*, Meltemi, Rome.

—(2005). *Semiotics Unbounded*, Toronto University Press, Toronto.

Ponzio, Augusto (1972). 'On Language as Work and Trade. Review article', *Semiotica* VI, 4, pp. 378-389.

— (1973). *Produzione linguistica e ideologia sociale*, De Donato, Bari; Fr. trans. *Production linguistique et idéologie sociale*, Les Editions Balzac,Candiac (Québec) 1992.

— (1986). 'On the Signs of Rossi-Landi's Work', *Semiotica* 62, 3/4, pp. 207-221.

— (1988). *Rossi-Landi e la filosofia del linguaggio*, Adriatica, Bari.

— (1989). 'Semiotics and Marxism', in T. A. Sebeok and J. Umiker-Sebeok (ed.), *The Semiotic Web*, 387-414, Mouton De Gruyter, Berlin.

— (1990). *Man as a Sign. Essays on the Philosophy of Language*, intro., trans., and ed. by S. Petrilli, Mouton, Berlin.

— (1991). *Filosofia del linguaggio 2. Segni valori ideologie*, Adriatica, Bari.

— (1993a). *Signs, Dialogue, and Ideology*, ed. by S. Petrilli, John Benjamins, Amsterdam.

— (1993b). *La ricerca semiotica* (with O. Calabrese and S. Petrilli), Esculapio, Bologna.

— (1994). *Fondamenti di filosofia del linguaggio* (con P. Calefato e S. Petrilli), Laterza, Bari 1999.

— (a cura di) (1996b). *Comunicazione, comunità, informazione,* Manni, Lecce.

— (1999). *La comunicazione*, Graphis, Bari.

Posner, R.; Robering, K.; Sebeok, T. A. (eds.) (1997, 1998, 2003). *Semiotik Semiotics. A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, 3 vols., Walter de Gruyter, Berlin.

Rossi-Landi, Ferruccio (1968). *Il linguaggio come lavoro e come mercato*, Bompiani, Milan 1992; Eng. trans. M. Adams *et alii*, *Language as Work and Trade,* Bergin and Garvey, South Had1ey (Mass.) 1983.

—(1972a). *Semiotica e ideologia*, Bompiani, Milan 1994.

—(1978a). *Ideologia,* ISEDI, Milan; Milan, 1982; new ed. with introduction by A. Ponzio, Rome/ Meltemi, 2003; Eng. trans. R. Griffin, *Marxism and Ideology*, Clarendon, Oxford 1990.

—(1992). *Between Signs and Non-signs*, intro. and ed. by S. Petrilli, John Benjamins, Amsterdam.

Sebeok, Thomas A. (ed.) (1968). *Animal Communication: Techniques of Studies and Results of Research*, Indiana University Press, Bloomington, IN.

—(1972). *Perspectives in Zoosemiotics*, Mouton, The Hague.

—(1976). *Contributions to the Doctrine of Signs*, Indiana University Press. Bloomington, IN; 2nd ed. University Press of America, Lanham 1979; It. trans. by M. Pesaresi, *Contributi alla dottrina dei segni*, Feltrinelli, Milan 1979.

—(ed.) (1978). *Sight, Sound and Sense*, Indiana University Press, Blooomington, IN.

—(ed.) (1979). *How Animals Communicate*, Indiana University Press, Bloomington, IN.

—(1979). *The Sign & Its Masters*, The University of Texas Press, Texas; 2nd ed. University Press of America, Lanham, Maryland; It. trans. and intro. by S. Petrilli, *Il segno e i suoi maestri*, Adriatica, Bari 1985.

—(1981). *The Play of Musement,* Indiana University Press, Bloomington, IN; It. trans. by M. Pesaresi, *Il gioco del fantasticare*, Spirali, Milan 1984.

—(1986). *I Think I Am a Verb*, Plenum Press, New York; It. trans. and intro. by S. Petrilli, *Penso di essere un verbo*, Sellerio, Palermo 1990.

—(1990). *Essays in Zoosemiotics*, ed. by M. Danesi, University of Toronto Press, Toronto.

—(1991a). *American Signatures. Semiotic Inquiry and Method*, ed. by I. Smith, Oklahoma Press, Norman-London.

—(1991b). *A Sign Is Just a Sign*, Indiana University Press, Bloomington, IN; It. trans. and intro. by S. Petrilli, *A sign is just a sign. La semiotica globale,* Spirali, Milan 1998.

—(1991c). *Semiotics in the United States*, Indiana University Press, Bloomington, IN; It. trans. intro. and ed. by S. Petrilli, *Sguardo sulla semiotica americana*, Bompiani, Milan 1992.

—(1994a). 'Global Semiotics', in I. Rauch and G. F. Carr, *Semiotics Around the World: Synthesis in Diversity*, Proceedings of the Vth International Congress of the International Association for Semiotic Studies, June 18, 1994, University of California, Berkeley, 105-130; Mouton de Gruyter, Berlin; now in T. A. Sebeok 2001; It. trans. by S. Petrilli, 'La semiotica globale', in T. A. Sebeok 1991b, it. trans., and in Sebeok 1997.

—(1994b). *Signs. An Introduction to Semiotics*, Toronto University Press, Toronto 2001.

—(1997). 'La semiotica globale', It. tran. by S. Petrilli, *Symbolon* 1, 1-2, 11-56.

—(1998). *Come comunicano gli animali che non parlano*, It. trans., intro. and ed. by S. Petrilli, Edizioni dal Sud, Bari.

—(2001a). *Global Semiotics*, Indiana University Press, Bloomington, IN.

—(2001b). 'Nonverbal communication', in P. Cobley (ed.), *Semiotics and Linguistics*, The Routledge Companion, London, pp. 14-27.

—(2001c). *The Swiss Pioneer in Nonverbal Communication Studies. Heini Hediger*, Legas, Toronto.

Sebeok, Thomas A.; Danesi, Marcel (2000). *The Forms of Meaning. Modeling Systems Theory and Semiotic Analysis*, Mouton de Gruyer, Berlin.

Sebeok, Thomas A.; Petrilli, Susan (1998). 'Women in Semiotics', in G. F. Carr *et alii* (ed. by), *Interdigitations: Essays for Irmengard Rauch*, Thoemmes Publisher, Bristol.

Sebeok, Thomas A.; Petrilli, Susan; Ponzio, Augusto (2001). *Semiotica dell'io*, Meltemi, Rome.

Uexküll, Jakob von (1909). *Umwelt und Innenwelt der Tiere*, Springer Verlag, Berlin.

— (1992). *Jakob von Uexküll's 'A Stroll through the Worlds of Animals and Men',* ed. Thure von Uexküll. *Semiotica*, Special Issue 89, 4.

Creative Multimodal Abduction
Neural Processes, External Semiotic Anchors, Hybrid Representations

Lorenzo Magnani
Department of Philosophy and Computational Philosophy Lab
University of Pavia
lmagnani@unipv.it

1. The Centrality of Abduction

If we decide to increase knowledge on both cognitive and semiotic aspects of hypothetical thinking it is necessary to develop a cognitive model of creativity able to represent not only "novelty" and "unconventionality", but also some features commonly referred to as the entire creative process, such as the hybrid modeling activity developed in the interplay between internal and external representations. The philosophical concept of *abduction* may be a candidate to solve this problem, and offers an approach to model creative processes of meaning generation in a completely explicit and formal way, which can fruitfully integrate the narrowness proper of a merely psychological approach, too experimentally human-oriented.

A hundred years ago, C. S. Peirce (Peirce 1931-1958) coined the concept of abduction in order to illustrate that the process of scientific discovery is not irrational and that a methodology of discovery is possible. Peirce interpreted abduction essentially as an "inferential" *creative process* of generating a new hypothesis. Abduction has a logical form – fallacious, if we model abduction by using classical syllogistic logic – distinct from deduction and induction. Reasoning which starts from reasons and looks for consequences is called *deduction*; that which starts from consequences and looks for reasons is called *abduction*.

Abduction – a distinct form of reasoning – 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 a): 1) abduction that only generates "plausible" hypotheses ("selective" or "creative") and 2) abduction considered as inference "to the best explanation", which also evaluates hypotheses (cf. Figure 1). An illustration from the field of medical knowledge is represented by the discovery of a new disease and the manifestations it causes which 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.

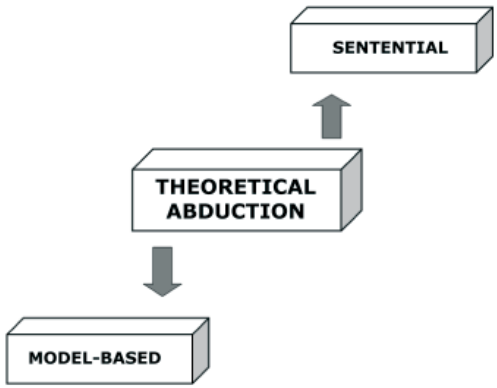


Figure 1. Creative and selective abduction.

In (Magnani 2001 a) I have introduced the concept of *theoretical abduction* as a form of neural and basically internal processing. I maintain 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, section 3 and subsection 2 (cf. Figure 2).

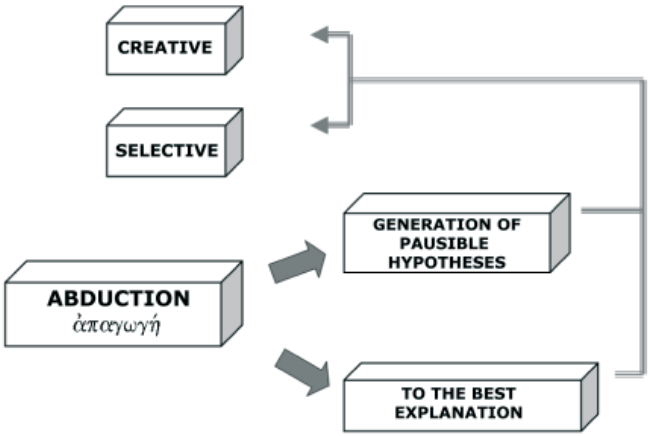


Figure 2. Theoretical abduction.

Theoretical abduction 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 I have called *epistemic mediators* (Magnani 2001 a). The concept of *manipulative abduction*[1] captures a large part of scientific 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 Manipulative abduction and epistemic mediators are introduced and illustrated in (Magnani 2001 a) and (Magnani 2001 b).

In section 3 I will describe how manipulative abduction can nicely account for the relationship between meaningful behavior and dynamical interactions with the environment. The following sections illustrate that at the roots of the creation of new meanings there is a process of externalization that exhibits a new cognitive description of the mechanisms underling the emergence of meaning processes through semiotic delegations to the environment.

2. Model-Based Abduction and Semiosis beyond Peirce

What exactly is model-based abduction from a philosophical point of view? Peirce stated that all thinking is in signs, and signs can be icons, indices, or symbols and that all *inference* is a form of sign activity, where the word sign includes "feeling, image, conception, and other representation" (Peirce 1931-1958, 5.283). In this light it can be maintained that a considerable part of the creative meaning processes is *model-based.* Moreover, a considerable part of meaning creation processes (not only in science) occurs in the middle of a relationship between brains and external objects and tools that have received cognitive and/or epistemological delegations (cf. the following section).

Recognizing in a series of *material*, physical events, that they make up a series of signs, is to know the existence of a "mind" (or of a group of minds) and to be absorbed in making, manifesting, or reacting to a series of signs is to be absorbed in, Peirce says, "being a mind". "[. . .] all thinking is dialogic in form" (Peirce 1931-1958, 6.338), both at the intrasubjective and intersubjective level, so that we see ourselves exactly as others see us, or see them exactly as they see themselves, and we see ourselves through our own speech and other interpretable behaviors, just others see us and themselves in the same way, in the commonality of the whole process (Brent 2000, p. 10).

As I will better explain later on in the following sections, in this perspective minds are material like brains, in so far as they consist in intertwined internal and external semiotic processes; Peirce nicely observes: "[. . .] the psychologists undertake to locate various mental powers in the brain; and above all consider it as quite certain that the faculty of language resides in a

certain lobe; but I believe it comes decidedly nearer the truth (though not really true) that language resides in the tongue. In my opinion it is much more true that the thoughts of a living writer are in any printed copy of his book than they are in his brain" (Peirce 1931-1958, 7.364].

3. Constructing Meaning through Mimetic and Creative External Objects

3.1 Manipulating Meanings through External Semiotic Anchors

Manipulative abduction occurs when many external things, usually inert from the semiotic point of view, can be transformed into what I have called, in the case of scientific reasoning, "epistemic mediators" (Magnani 2001 a) that give rise to new signs, new chances for what Peirce and semiotics calls "interpretants", and new interpretations.

We can cognitively account for this process of externalization[2] taking advantage of the concept of *manipulative* abduction. It happens when we are thinking *through* doing and not only, in a pragmatic sense, about doing. For instance, when we are creating geometry constructing and manipulating an external suitably constructed icon like a triangle looking for new meaningful features of it, like in the case given by Kant in the "Transcendental Doctrine of Method" ((Magnani 2001 b) and the following subsection). It 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 (semantic) practices.

If the structures of the environment play such an important role in shaping our semiotic representations and, hence, our cognitive processes, we can expect that physical manipulations of the environment receive a great cognitive relevance.

Several authors have pointed out the role that physical actions can have at a cognitive level. In this sense Kirsh and Maglio (Kirsh and Maglio 1994)

2 A significant contribution to the comprehension of this process in terms of the so–called "disembodiment of the mind" derives from some studies in the field of cognitive paleoanthropology. In (Magnani 2006) I have illustrated this issue relating it to the Turing ideas on "unorganized" and "organized" brains.

distinguish actions into two categories, namely *pragmatic actions* and *epistemic actions*. Pragmatic actions are the actions that an agent performs in the environment in order to bring itself physically closer to a goal. In this case the action modifies the environment so that the latter acquires a configuration that helps the agent to reach a goal which is understood as physical, that is, as a desired state of affairs. Epistemic actions are the actions that an agent performs in a semiotic environment in order to discharge the mind of a cognitive load or to extract information that is hidden or that would be very hard to obtain only by internal computation.

In this section I want to focus specifically on the relationship that can exist between manipulations of the environment and representations. In particular, I want to examine whether external manipulations can be considered as means to construct external representations.

If a manipulative action performed upon the environment is devoted to create a configuration of signs that carries relevant information, that action will well be able to be considered as a cognitive semiotic process and the configuration of elements it creates will well be able to be considered an external representation. In this case, we can really speak of an embodied cognitive process in which an action constructs an external representation by means of manipulation. We define *cognitive manipulating* as any manipulation of the environment devoted to construct external configurations that can count as representations.

An example of cognitive manipulating is the diagrammatic demonstration illustrated in Figure 3, taken from the field of geometry. In this case a simple manipulation of the triangle in Figure 3(a) gives rise to an external configuration -Figure 3(b) -that carries relevant semiotic information about the internal angles of a triangle "anchoring" new meanings.

The entire process through which an agent arrives at a physical action that can count as cognitive manipulating can be understood by means of the concept of manipulative abduction (Magnani 2001 a). Manipulative abduction is a specific case of cognitive manipulating in which an agent, when faced with an external situation from which it is hard or impossible to extract new meaningful features of an object, selects or creates an action that struc-

tures the environment in such a way that it gives information which would be otherwise unavailable and which is used specifically to infer explanatory hypotheses.

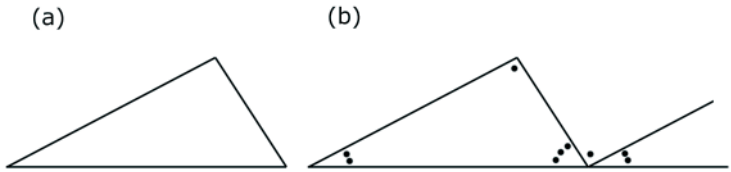


Figure 3. Diagrammatic demonstration that the sum of the internal angles of any triangle is 180◦ . (a) Triangle. (b) Diagrammatic manipulation/construction.

In this way the semiotic result is achieved on *external* representations used in lieu of the internal ones. Here action performs an *epistemic* and not a merely performatory role, for example relevant to abductive reasoning.

3.2 Geometrical Construction is a Kind of Manipulative Abduction

Let's quote Peirce's passage about mathematical constructions. Peirce says that mathematical and geometrical reasoning "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. All valid necessary reasoning is in fact thus diagrammatic" (Peirce 1931-1958, 1.54). This passage clearly refers to a situation like the one I have illustrated in the previous subsection. This kind of reasoning is also called by Peirce "theorematic" and it is a kind of "deduction" necessary to derive significant theorems: "[. . .] is one which, having represented the conditions of the conclusion in a diagram, performs an ingenious experiment upon the diagram, and by observation of the diagram, so modified, ascertains the truth of the conclusion" (Peirce 1931-1958, 2.267). The experiment is performed with the help of "imagination upon the image of the premises in order from the result of such experiment to make corollarial deductions to the truth of the conclusion" (Peirce 1976, IV, p. 38). The "corollarial" reasoning is mechanical (Peirce thinks it can be performed by a "logical machine") and not creative, "A Corollarial Deduction is

one which represents the condition of the conclusion in a diagram and finds from the observation of this diagram, as it is, the truth of the conclusion" (Peirce 1931-1958, 2.267) (cf. also Hoffmann 1999).

In summary, the point of theorematic reasoning is the transformation of the problem by establishing an unnoticed point of view to get interesting – and possibly new – insights. The demonstrations of theorems in mathematics are examples of theorematic deduction.

Not dissimilarly Kant says that in geometrical construction of external diagrams "[. . .] I must not restrict my attention to what I am actually thinking in my concept of a triangle (this is nothing more than the mere definition); I must pass beyond it to properties which are not contained in this concept, but yet belong to it" (Kant 1787, A718-B746, p. 580).

We have seen that manipulative abduction is a kind of abduction, usually model-based, that exploits external models endowed with delegated (and often implicit) cognitive and semiotic roles and attributes.

- The model (diagram) is *external* and the strategy that organizes the manipulations is unknown *a priori*.
- The result achieved is *new* (if we, for instance, refer to the constructions of the first creators of geometry), and adds properties not contained before in the concept (the Kantian to "pass beyond" or "advance beyond" the given concept (Kant 1787, A154-B193/194, p. 192]).[3]

Iconicity in theorematic reasoning is central. Peirce, analogously to Kant, maintains that "philosophical reasoning is reasoning with words; while theorematic reasoning, or mathematical reasoning is reasoning with specially constructed schemata" (Peirce 1931-1958, 4.233); moreover, he uses diagrammatic and schematic as synonyms, thus relating his considerations

3 Of course in the case we are using diagrams to demonstrate already known theorems (for instance in didactic settings), the strategy of manipulations is not necessary unknown and the result is not new.

to the Kantian tradition where schemata mediate between intellect and phenomena.[4]

As we have already said, for Peirce the whole mathematics consists in building diagrams that are "[. . .] (continuous in geometry and arrays of repeated signs/letters in algebra) according to general precepts and then [in] observing in the parts of these diagrams relations not explicitly required in the precepts" (Peirce 1931-1958, 1.54). Peirce contends that this diagrammatic nature is not clear if we only consider syllogistic reasoning "which may be produced by a machine" but becomes extremely clear in the case of the "logic of relatives, where any premise whatever will yield an endless series of conclusions, and attention has to be directed to the particular kind of conclusion desired" (Peirce 1987, pp. 11–23).

In ordinary geometrical proofs auxiliary constructions are present in terms of "conveniently chosen" figures and diagrams where strategic moves are important aspects of deduction. The system of reasoning exhibits a dual character: deductive and "hypothetical". Also in other – for example logical -deductive frameworks there is room for strategical moves which play a fundamental role in the generations of proofs. These strategical moves correspond to particular forms of abductive reasoning.

We know that the kind of reasoned inference that is involved in creative abduction goes beyond the mere relationship that there is between premises and conclusions in valid deductions, where the truth of the premises guarantees the truth of the conclusions, but also beyond the relationship that there is in probabilistic reasoning, which renders the conclusion just more or less probable. On the contrary, we have to see creative abduction as formed by the application of *heuristic procedures* that involve all kinds of good and bad inferential actions, and not only the mechanical application of rules. It is only by means of these heuristic procedures that the acquisition of *new* truths is guaranteed. Also Peirce's mature view illustrated above on creative

4 Schematism, a fruit of the imagination is, according to Kant, "[. . .] 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 1787, A141-B181, p. 183)

abduction as a kind of inference seems to stress the strategic component of reasoning.

The example of diagrams in geometry furnishes a semiotic and epistemological example of the nature of the cognitive interplay between internal neuronal representations (and embodied "cognitive" kinesthetic abilities) and external representations I have illustrated above: also for Peirce, more than a century before the new ideas derived from the field of distributed reasoning, the two aspects are intertwined in the pragmatic and semiotic view, going beyond the rigidity of the Kantian approach in terms of schematism. Diagrams are icons that take material and semiotic form in an external environment endowed with

- constraints depending on the specific cognitive delegation performed by human beings
  and
- the particular intrinsic constraints of the materiality at play.

Concrete manipulations on them can be done for instance to get new data and cognitive information and/or to simplify the problem at issue.

3.3 The Semiosis of Re-Embodiment

Some interesting semiotic aspects of the above illustrated process can be nicely analyzed. Imagine that a suitable *fixed internal record* exists – deriving from the cognitive exploitation of the previous suitable interplay with *external structures* -at the level of neural activation and that for instance it embeds an abstract concept endowed with all its features, for example the concept of triangle. Now, the human agent, via neural mechanisms and bodily actions, can "reembody" that concept by making an external perceivable *sign*, for instance available to the attention of other human or animal senses and brains. For instance that human agent can use what in semiotics is called a *symbol* (with its conventional character: *ABC*, for example), but also an *icon* of relations (a suitable diagram of a triangle), or a *hybrid representation* that will take advantage of both. In Peircian terms:

> A representation of an idea is nothing but a sign that calls up another idea. When one mind desires to communicate an idea to another, he embodies his idea by making an outward perceptible image which directly calls up a like idea; and another mind perceiving that image gets a like idea. Two persons may agree upon a conventional sign which shall call up to them an idea it would not call up to anybody else. But in framing the convention they must have resorted to the primitive diagrammatic method of embodying the idea in an outward form, a picture. Remembering what likeness consists in, namely, in the natural attraction of ideas apart from habitual outward associations, I call those signs which stand for their likeness to them *icons*. Accordingly, I say that the only way of directly communicating an idea is by mean of an icon; and every indirect method of communicating an idea must depend for its establishment upon the use of an icon (Peirce 1967, 787, pp. 26–28).

It is well-known that for Peirce every picture is a icon and thus every diagram, even if it lacks a sensuous similarity with the object, but just exhibits an analogy between the relations of the part of it and of the object: Particularly deserving of notice are icons in which the likeness is aided by conventional rules. Thus, an algebraic formula is an icon, rendered such by the rules of commutation, association, and distribution of the symbols; that it might as well, or better, be regarded as a compound conventional sign. It may seem at first glance that it is an arbitrary classification to call an algebraic expression an icon; that it might as well, or better, be regarded as a compound of conventional sign. But it is not so. For a great distinguishing property of the icon is that by direct observation of it other truths concerning its object can de discovered than those which suffice to determine its construction. Thus, by means of two photographs a map can be drawn, etc. Given a conventional or other general sign of an object, to deduce any other truth than which it explicitly signifies, it is necessary, in all cases, to replace that sign by an icon. This capacity of revealing unexpected truth is precisely that wherein the utility of algebraic formulae consists, so that the icon in character is the prevailing one (Peirce 1967, 787, pp. 26–28).

Stressing the role of iconic dimensions of semiosis[5] in the meantime celebrates the virtues of analogy, as a kind of "association by resemblance", as contrasted to "association by contiguity"

Human beings delegate cognitive features to external representations through semiotic attributions because for example in many problem solving situations the internal computation would be impossible or it would involve a very great effort because of human mind's limited capacity. As I will illustrate below in this paper I think there are two kinds of artifacts that play the role of *external objects* (representations) active in this process of externalization of the mind: *creative* and *mimetic*. Mimetic external representations mirror concepts and problems that are already represented in the brain and need to be enhanced, solved, further complicated, etc. so they sometimes can creatively give rise to new concepts and meanings.

Following my perspective it is at this point evident that the "mind" transcends the boundary of the individual and includes parts of that individual's environment. It is in this sense that the mind is semiotic and artificial.

3.4 External Diagrammatization and Iconic Brain Co-Evolution

Human brains organize themselves through a semiotic activity that is reified in the external environment and then re-projected and reinterpreted through new configurations of neural networks and chemical processes. I have illustrated in a previous paper (Magnani 2006) that through the mediation of the material culture the modern human mind for example can arrive to *internally* "think" the new complicated meaning of animals and people at the same time. We can account for this process of externalization from an impressive cognitive point of view.

I maintain that representations are external and internal. Following our previous considerations it would seem that diagrams can be fruitfully seen from a semiotic perspective as external representations expressed through icons and symbols, aimed at simply "mimicking" various humans' internal images.

5 We have to remember that in this perspective any proposition is a diagram as well, because it represents a certain relation of symbols and indices.

I think there are two basic kinds of external representations active in this process of externalization of the mind: *creative* and *mimetic*. Mimetic external representations mirror concepts and problems that are already represented in the brain and need to be enhanced, solved, further complicated, etc. so they sometimes can creatively give rise to new concepts and meanings. In our geometrical example it is clear how for instance a mimetic geometric representation can become creative and give rise to new meanings and ideas in the hybrid interplay between brains and suitable "cognitive niches"[6] that consequently are appropriately reshaped. Hence representations can also play the role of creative representations human beings externalize and manipulate not just to mirror the internal ways of thinking of human agents but to find room for concepts and new ways of inferring which cannot - at a certain time - be found internally "in the mind".

In summary, we can say that

- diagrams as *external* iconic (often enriched by symbols) representations are formed by external materials that either mimic (through reification) concepts and problems already internally present in the brain or creatively express concepts and problems that do not have a semiotic "natural home" in the brain;
- subsequent *internalized diagrammatic* representations are internal reprojections, a kind of recapitulations, (learning) in terms of neural patterns of activation in the brain ("thoughts", in Peircian sense), of external diagrammatic representations. In some simple cases complex diagrammatic transformations – can be "internally" manipulated *like* external objects and can further originate new internal reconstructed representations through the neural activity of transformation and integration.

6 This expression, used in the different framework of the problem of language as biological adaptation to the environment appears very appropriate also in this context (Pinker 2003).

This process explains – from a cognitive point of view – why human agents seem to perform both computations of a connectionist type[7] such as the ones involving representations as

- (I Level) patterns of neural activation that arise as the result of the interaction (also presemiotic) between body and environment (and suitably shaped by the evolution and the individual history): pattern completion or image recognition, and computations that use representations as
- (II Level) derived combinatorial syntax and semantics dynamically shaped by the various artificial external representations and reasoning devices found or constructed in the semiotic environment (for example iconic representations); they are – more or less completely – neurologically represented contingently as patterns of neural activations that "sometimes" tend to become stabilized meaning structures and to fix and so to permanently belong to the I Level above.

It is in this sense we can say the "System of Diagrammatization", in Peircian words, allows for a self-controlled process of thought in the fixation of originally vague beliefs: as a system of learning, it is a process that leads from "absolutely undefined and unlimited possibility" (Peirce 1931-1958, 6.217) to a fixation of belief and "by means of which any course of thought can be represented with exactitude" (Peirce 1931-1958, 4.530). Moreover, it is a system which could also improve other areas of science, beyond mathematics, like logic, it "greatly facilitates the solution of problems of Logic. [. . .] If logicians would only embrace this method, we should no longer see attempts to base their science on the fragile foundations of metaphysics or a psychology not based on logical theory" (Peirce 1931-1958, 4.571).

7 Here the reference to the word "connectionism" is used on the plausible assumption that all mental representations are brain structures: verbal and the full range of sensory representations are neural structures endowed with their chemical functioning (neurotransmitters and hormones) and electrical activity (neurons fire and provide electrical inputs to other neurons). In this sense we can reconceptualize cognition neurologically: for example the solution of a problem can be seen as a process in which one neural structure representing an explanatory target generates another neural structure that constitutes a hypothesis for the solution.

The I Level originates those sensations (they constitute a kind of "face" we think the world has), that provide room for the II Level to reflect the structure of the environment, and, most important, that can follow the computations suggested by the iconic external structures available. It is clear that in this case we can conclude that the growth of the brain and especially the synaptic and dendritic growth are profoundly determined by the environment. Consequently we can hypothesize a form of co-evolution between what we can call the *iconic brain* and the development of the external diagrammatic systems. Brains build iconic signs as diagrams in the external environment learning from them new meanings through interpretation (both at the spatial and sentential level) after having manipulated them.

When the fixation is reached – imagine for instance the example above, that fixes the sum of the internal angles of the triangle – the pattern of neural activation no longer needs a direct stimulus from the external spatial representation in the environment for its construction and can activate a "final logical interpretant", in Peircian terms. It can be neurologically viewed as a fixed internal record of an external structure (a fixed belief in Peircian terms) that can exist also in the absence of such external structure. The pattern of neural activation that constitutes the I Level Representation has kept record of the experience that generated it and, thus, carries the II Level Representation associated to it, even if in a different form, the form of *semiotic memory* and not the form of the vivid *sensorial experience* for example of the triangular construction drawn externally, over there, for instance in a blackboard. Now, the human agent, via neural mechanisms, can retrieve that II Level Representation and use it as an internal representation (and can use it to construct new internal representations less complicated than the ones previously available and stored in memory) (Gatti and Magnani 2006).

At this point we can easily understand the particular *mimetic* and *creative* role played by external diagrammatic representations in mathematics: some concepts, meanings, and "ways of [geometrical] inferring" performed by the biological human agents appear hidden and tacit and can be rendered explicit by building external diagrammatic mimetic models and structures; later on the agent will be able to pick up and use what was suggested by the

constraints and features intrinsic and immanent to their external semiotic materiality and the relative established conventionality: artificial languages, proofs, new figures, examples, etc.;

some concepts, meanings, and "new ways of inferring" can be discovered only through a problem solving process occurring in a distributed interplay between brains and external representations. I have called this process externalization (or disembodiment) of the mind: the representations are mediators of results obtained and allow human beings

(a) to re-represent in their brains new concepts, meanings, and reasoning devices picked up outside, externally, previously absent at the internal level and thus impossible: first, a kind of alienation is performed, second, a recapitulation is accomplished at the neuronal level by re-representing internally that which has been "discovered" outside. We perform cognitive geometric operations on the structure of data that synaptic patterns have "picked up" in an analogical way from the explicit diagrammatic representations in the environment;

(b) to re-represent in their brains portions of concepts, meanings, and reasoning devices which, insofar as explicit, can facilitate inferences that previously involved a very great effort because of human brain's limited capacity. In this case the thinking performance is not completely processed internally but in a hybrid interplay between internal (both tacit and explicit) and external iconic representations. In some cases this interaction is between the internal level and a computational tool which in turn can exploit geometrical representations to perform inferences (cf. above subsection 3.1).

An evolved mind is unlikely to have a natural home for complicated creative concepts like the ones geometry introduced, as such concepts do not exist in a definite way in the natural (not artificially manipulated) world: so whereas evolved minds could construct spatial frameworks and perform some trivial spatial inferences in a more or less tacit way by exploiting modules shaped by natural selection, how could one think exploiting explicit complicated geometrical concepts without having picked them up outside, after having produced them?

A mind consisting of different separated implicit templates of thinking and modes of inferences exemplified in various exemplars expressed through natural language cannot come up with certain mathematical and geometrical entities without the help of the external representations. The only way is to extend the mind into the material world, exploiting paper, blackboards, symbols, artificial languages, and other various semiotic tools, to provide *semiotic anchors* for finding ways of inferring that have no natural home within the mind, that is for finding ways of inferring and concepts that take us beyond those that natural selection and cultural training could enable us to possess at a certain moment.

Hence, we can hypothesize – for example – that many valid spatial reasoning habits which in human agents are performed internally have a deep origin in the past experience lived in the interplay with iconic systems at first represented in the environment. As I have just illustrated other recorded thinking habits only partially occur internally because they are hybridized with the exploitation of already available or suitably constructed external diagrammatic artifacts.

3.5 Delegated and Intrinsic Constraints in External Agents

We have said that through the cognitive interplay with external representations the human agent is able to pick up and use what suggested by the constraints and features intrinsic to their external materiality and to their relative established conventionality: artificial languages, proofs, examples, etc. Let us consider the example above (section 3.1) of the sum of the internal angles of a triangle. The human agent

- looks, through diagram manipulations, for possible necessary consequences that occur over there, in the diagram/icon and that obey both
- the conventional *delegated* properties and
- the properties *intrinsic* to the materiality of the model.

This external model is a kind of autonomous cognitive *agent* offered to new Peircian interpretants of the problem/object in question. The model

can be picked up later and acknowledged by the human agent through fixation of a new neural configuration – a new "thought" (in the case the new result concerning the sum of the internal angles).

Let us imagine we choose a different representational system still exploiting material and external diagrams. Through the manipulation of the new symbols and diagrams we expect very different conclusions. An example is the one of the non-Euclidean discoveries. In Euclidean geometry, by adopting the postulate of parallels we necessarily arrive to the ineluctable conclusion that the sum of internal angles of a triangle is 180 ° , but this does not occur in the case of the non-Euclidean geometry where a different selected representational system -that still uses Euclidean icons -determines quite different possibilities of constructions, and thus different results from iconic experimenting.

4. Conclusion

The main thesis of this paper is that the process of externalization of mind is a significant cognitive perspective able to unveil some basic features of *creative* manipulative abductive reasoning. Its fertility in explaining the semiotic interplay between internal and external levels of cognition is evident. I maintain that various aspects of creative meaning formation could take advantage of the research on this interplay: for instance study on external mediators can provide a better understanding of the *hybrid* processes of explanation and discovery in science and in some areas of artificial intelligence related to mechanizing discovery processes.